Intrinsic vowel normalization: Comparing different procedures (data from Tuscan Italian)

Silvia Calamai

Several intrinsic methods used in vowel representation were evaluated; the performance of forty-two different combinations of ways to normalize vowels (i.e. different scales, *f*0-correction, formant-correction) was compared by means of Multivariate Analyses of Variance and Discriminant Analyses. Data came from two different Tuscan varieties (Pisa and Florence) and from two different speech styles (read and semi-spontaneous speech). Intrinsic vowel normalization procedures seemed effective at maximizing differences between vowel categories but seemed somehow deficient in minimization differences in the same vowel spoken by different speakers. Comparisons across different styles within the same speakers proved to be more successful than comparisons across different dialects and different speakers, although in the second case an identical type of speech material was used. As a whole, the parameter set F1 x F2 on the ERB scale seemed to be one of the most successful in reducing inter-speaker variability and in preserving vowel-category separability.*

0. Introduction

A single vowel sound can be depicted as an acoustic event or as an auditory event. From an acoustic point of view, it is usually characterised by its first two formants, F1 and F2 (the physical correlates of 'vowel height' and 'place of constriction', respectively).[1] The linguistic quality of a vowel is drawn by depicting a vowel as a point in a two-dimensional space, with F2 along the horizontal axis, and F1 along the vertical axis ('vowel diagram' or 'formant chart'). A convenient way of representing a given vowel system is given by the F2/F1 plane, in which a specific position is assumed by each member of a vowel system.

Acoustic analysis is unbiased and rapid; but if the output of the measurements has to be considered, it seems difficult to distinguish between relevant linguistic variation and irrelevant speaker-dependent variation. Different vowels for different speakers can fall at the same place in the F1-F2 plot; and the same vowels for different speakers can fall at different places in the F1-F2 plot. In other words, if different speakers pronounce the same vowel, their formants are often not in the exact same position in the vowel diagram. The

Rivista di Linguistica 17.2 (2005), pp. 211-270 *(ricevuto nel marzo 2005)*

reverse is also possible: two different vowels spoken by two different speakers may end up on the same spot in the diagram. In addition, the acoustic properties of vowels vary depending not only on the individual speaker,[2] but also depending on changes in stress, on the rate of speech, and on the phonetic context in which the vowel occurs. In general, the vowel systems of different speakers represented by measured formants cannot be superimposed.

Despite linguistic, paralinguistic and extralinguistic variation, listeners nevertheless show a great amount of perceptual constancy in recognizing vowels from different speakers under various speech conditions: how do they determine the phonetic quality of a vowel and abstract away from the idiosyncratic speaker characteristics? It has been supposed that listeners make use of some kind of normalization procedure which enables them to adjust their perceptual strategies.[3] Two general approaches aim at solving the problem here outlined: the first one deals with 'extrinsic specifications', such as vocal tract length, a speaker's entire vowel space, vowel formant range in a preceding context; and the second one deals with the so-called 'intrinsic methods'.[4] Vowel-intrinsic information (i.e. *f*0, higher formants, duration) are assumed to be all that is necessary to normalize vowels. Whereas in extrinsic specifications a frame of reference is assumed to be outside a single vowel token, intrinsic methods hypothesize that there is sufficient information within the acoustic pattern of the vowel itself to allow the listener to identify it unambiguously: intrinsic methods try to reduce speaker-specific variance by representing the vowels with different scales and/or with *f*0-correction or with formant correction. Intrinsic and extrinsic representations predict that normalization occurs at different moments in vowel processing: at a peripheral auditory stage in the former case, at a more central processing stage in the latter. As Nearey (1989:2088) pointed out, all the procedures outlined above are 'data analytic' rather than 'perceptual', since "they deal with reliable separation of categories based on data from production measurements only".

The present paper will concentrate on some of the most common intrinsic methods used in vowel classification. By means of statistical analyses (analytically described in 0.3) it intends to investigate which factors best distinguish and characterize a vowel system and which kind of procedures reduces the variety of vowel distribution. It therefore aims at acquiring more insight into the nature of speaker and vowel specific variance by evaluating the workings of different potentially normalizing factors.

0.1. Normalizing factors

Two types of normalizing factors will be evaluated in the following pages: scale and correction.

Scale. Formants are usually displayed in Hertz, which is the physical measure of cycles per second (frequency). But vowels can also be treated in perceptual terms: in this kind of representation the application of some auditory transform carrying physical frequency into pitch is required. Five auditory transforms are usually used to generate auditory vowel spaces: the logarithmic (Log) scale, the Koenig-scale, the Mel scale, the Bark scale, the ERB-rate scale. Auditory scales emphasize the lowest formants (i.e. the ones below 1000 Hz), which hold a relevant part in sounds perception; they incorporate one aspect of human hearing, specifically the fact that differences between lower frequencies are more noticeable than identical differences at higher frequencies.

In the Log transform, values in Hertz are displayed along a logarithmic scale.[5] The Koenig-scale is exactly linear below 1000 Hz and exactly logarithmic above 1000 Hz (Koenig 1949); it has been used in the 1952 pioneer study by Peterson and Barney on North-American Vowels.

The Mel scale is based on a subjective measure of pitch 'magnitude', and it is approximately linear below about 1000 Hz and approximately logarithmic above 1000 Hz.[6] The Bark transform maps acoustic measurements onto a psychophysical space: it is derived from measurements of the frequency selectivity of the human auditory system, as measured by the so-called critical bandwidth. The Bark scale is approximately linear below about 500 Hz and approximately logarithmic above 500 Hz.[7] The ERB-rate scale is a new variety of the Bark scale: the critical bands are narrower, especially at lower frequencies, than on the Bark scale; for frequencies below 500 Hz, the ERB-rate is neither linear, such as the Bark scale, nor logarithmic, but something in between. In the present study, the formula described in Hermes & van Gestel (1991:97)[8] was used.

Correction. As already mentioned, there are two types of correction: *f*0-correction and formant correction.

Source-filter theory assumes that source and filter are independent, but this assumption is not entirely correct. Various perceptual experiments have shown the influence of *f*0 on vowel quality: *f*0 can behave as a normalizing factor of formant displacements by influenc-

ing the perception of vowel quality,[9] and could therefore have a positive effect on variance reduction when used on data of group of speakers with different anatomical and physiological characteristics, like female, male and children. According to Hirahara & Kato (1992: 92), the use of *f*0 in vowel normalization seems more in line with an aspect of human perception than the other vowel normalization procedures: the ones based on formant frequencies may provide a technical solution, but they do not take into account human auditory processes.

Reasonably, *f*0-correction will not deeply improve the classification of vowels in the present case, since the subjects used in the analyses are all males of (more or less) the same age.[10] In any case, the parameters (F1-*f*0) and (F2-*f*0) were tested in the following statistical analyses.

Formant correction means subtracting the value of one formant from another one: the difference between second and first formant and the difference between third and second formant were used in the following statistical analyses.[11] It has been noted that formant frequency distances incorporate some sort of speaker normalization, since they are relatively constant across speakers (Traunmüller 1981; Syrdal & Gopal 1986). The 'effective second formant' (F2') – which is a weighted combination of F2 and higher formants – was not used in the present study. At any rate, the results of Hillenbrand & Gayvert (1993) did not show any improvement in category separability when F2' was substituted in place of F2.

Combination of different factors. Some of the vowel formant normalization algorithms are thus based on warpings of F1/F2 space using frequencies of higher formants, on the relationship between formants and *f*0, or on logarithmic transforms of values in linear Hertz; others (Bark, Mel, ERB) use psychoperceptual criteria. In the following pages, different scales and different kinds of correction were used in order to check the fourty-two combinations chosen.

Duration. Vowel duration is a temporal variable that could affect vowel classification, even in languages such as Italian where vowel duration by itself is not a phonological marker: close vowels tend to be briefer than open vowels, since the jaw is a sluggish articulator. In all the fourty-two combinations duration was added to the spectrally based measurements in order to verify its value for improving vowel categorization.

The normalization procedures tested in the present paper are displayed in Table 1. They are supposed to:

(i) maximize differences between vowel categories (that is to say, maximization of vowel-category separability);

(ii) minimize differences in the same vowel spoken by different talkers, particularly those differences associated with vocal-tract length variability (that is to say, minimization of within-vowel category differences among talkers).

Table 1. Overview of the transformations and normalization procedures used.

Scale	Formant frequencies	*f*0-correction	Formant-correction	*f*0-correction + formant-correction
Hertz	F1 x F2	F1-*f*0 x F2 F1-*f*0 x F2-*f*0	F1 x F2-F1 F2-F1 x F3-F2	F1-*f*0 x F2-F1 F1-*f*0 x F3-F2
Log(Hz)	F1 x F2	F1-*f*0 x F2 F1-*f*0 x F2-*f*0	F1 x F2-F1 F2-F1 x F3-F2	F1-*f*0 x F2-F1 F1-*f*0 x F3-F2
Koenig-scale	F1 x F2	F1-*f*0 x F2 F1-*f*0 x F2-*f*0	F1 x F2-F1 F2-F1 x F3-F2	F1-*f*0 x F2-F1 F1-*f*0 x F3-F2
Mel	F1 x F2	F1-*f*0 x F2 F1-*f*0 x F2-*f*0	F1 x F2-F1 F2-F1 x F3-F2	F1-*f*0 x F2-F1 F1-*f*0 x F3-F2
Bark	F1 x F2	F1-*f*0 x F2 F1-*f*0 x F2-*f*0	F1 x F2-F1 F2-F1 x F3-F2	F1-*f*0 x F2-F1 F1-*f*0 x F3-F2
ERB	F1 x F2	F1-*f*0 x F2 F1-*f*0 x F2-*f*0	F1 x F2-F1 F2-F1 x F3-F2	F1-*f*0 x F2-F1 F1-*f*0 x F3-F2

If an appropriate normalization procedure succeeds in removing the puzzling elements introduced by the heterogeneity of speakers, the resulting acoustic vowel chart becomes an accurate representation of the linguistic aspects of the vowels and facilitates both across-speaker and across-language comparison.

0.2. Speech material and plan of the paper

The data consisted of different speech material:

- isolated words and pseudo-words read once by six male subjects from Pisa (undergraduate students), all native speakers aged 23-31; and by two male subjects from Florence, both native speakers aged 25 and 26;

- connected speech material (quasi-spontaneous dialogues staged 'map tasks'), uttered by two speakers from Pisa (the same who read the word list).

The recordings were made in a sound-proof room at the Laboratorio di Linguistica of the Scuola Normale Superiore (Pisa).[12]

The seven stressed vowels (/i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/) and the five unstressed vowels (/i/, /e/, /a/, /o/, /u/) were used in the analyses; the diphthongs were excluded. Pisa and Florence speakers share the same phonological vowel system but they show a great amount of phonetic variation:[13] as for as certain vowel categories are concerned, the Pisa variety has (very) different acoustic targets (see *infra*).

The whole Pisa read speech sample was first observed (both in stressed and unstressed condition): the validity of all the normalization models tested here were therefore focussed on idiosyncratic effects, as the sex was the same and the age limited to a small range (see Part I). Register variation was then considered: a sub-*corpus* of the Pisa read speech material was compared with a *corpus* of semi-spontaneous speech material, in order to verify which vowel categories are better distinguished in the read speech and in the semi-spontaneous speech condition, and which are the best normalization methods for dealing with different speech styles (see Part II § 1).

The two groups of speakers – from Pisa and from Florence – were finally used in order to investigate some kind of geographical influence: this comparison allows to observe how well the normalization procedures maintained vowel identity and how well they revealed differences among vowel targets (see Part II § 2). In other words, our purpose was to investigate whether the results of the normalization described for the Pisa speakers were more or less the same as those obtained for the Florence speakers, apart from some particular differences concerning local features, like the lowering of /ɛ/ and /ɔ/ in the Pisa variety. Compared to the other Tuscan varieties, the Pisa Italian[14] shows two peculiar phonetic features: a pair of low vowels, both at the front and at the back side, and the shifting of /a/ into [ɑ]. In particular, the lowering of /ɛ/ is very easily perceived and it is a sort of social marker signalling community identity (Calamai 2002).

As for the Pisa sample, 2708 tokens total (1509 for the stressed vowels, 1199 for the unstressed vowels) were measured in the read speech condition. As for the comparison with the connected speech condition, a sub-sample of the Pisa stressed vowel in the read speech condition was used (360 tokens for the stressed vowels in the con-

nected speech material and 315 tokens for the stressed vowels in the read speech condition). As for the geographical comparison in the read speech condition, a sub-sample of the Pisa stressed vowel (315 tokens) was compared with 323 tokens from the Florence speech sample.

Using the Multispeech software, the speech material was down-sampled at 11025 Hz, and spectrographic analysis was performed using an LPC algorithm, Hamming window, and a frame length of 20 ms. Digital spectrograms were derived, using a pre-emphasis coefficient of 0.9; 14th order linear prediction was used for all data.[15] Where any one of the formants was incorrectly tracked, Long Term Average Power Spectrum, using the Fast Fourier Transform, was performed. The pitch analysis was made with the autocorrelation method (70-350 analysis range). Vowel onset and offset were determined by observing both the spectrogram and the waveform;[16] measures of F1, F2 and F3 were taken at vowel midpoint over three consecutive frames whose values were averaged; the measurements were all checked by visual inspection of linear-prediction-based formant tracks overlaid on spectrograms.

In short, the design of the experiments was therefore divided into two parts. In the first part, the whole Pisa sample was observed both in the stressed and in the unstressed condition. In the second part, a narrower sample was observed from two different points of view (speech style and language variety), as displayed in Table 2.

Table 2. Design of the Part II.

Variables	Group I	Group II
Speech styles	Read speech - Pisa sample	Semi-spontaneous speech - Pisa sample
Language variety	Read speech - Pisa variety	Read speech - Florence variety

0.3. Statistical evaluation

Following Adank (1999), in order to verify which normalization procedure was best at reducing variability while maintaining the identifiability of the vowel categories, 420 multivariate analyses of variance (MANOVA) and 504 discriminant analyses were performed on the whole Pisa sample (stressed and unstressed vowel system), on the two speech styles (read and semi-spontaneous speech) and on the two varieties (Pisa and Florence).

In each MANOVA, Hotelling's trace was extracted as a measure of variance reduction: this value represents the ratio between systematic variation and residual variation, or, in other words, the between ~ within variance ratio (the larger this ratio the more successful the transformation was in reducing the natural differences between speakers and their individual variability).[17]

Within each variety and within each speech style the MANOVAs were performed on the data of the seven stressed vowels, and they were performed on the data of the five unstressed vowels as well, as far as the whole Pisa sample is concerned. After performing the analyses, the scores were ranked and those with high F values got high rank scores[18] (I.1.1, I.2.1, II.1.1.1, II.1.2.1, II.2.1.1).

Discriminant analysis is a statistical technique which is relevant to the separation of types: its goal is to find a linear combination of the original parameters that provides the optimal separation between groups of points defined by the different types. Vowels are considered as *a priori* categories and the formant values (on different scales and with various kinds of correction) as discriminant variables.[19] Once the discriminant functions were computed, a recognition task was simulated in each sub-sample; from the linear discriminant analysis for the seven vowel categories, confusion matrices showing how the vowel tokens were classified were obtained. The percentage of correctly classified vowels was used as the criterion of success (I.1.2, I.2.2, II.1.1.2, II.1.2.2, II.2.1.2). Confusion matrices scores are related to the degree of overlap between groups (i.e. vowel categories) in a space defined by acoustic measurements: data showing a small degree of overlap possess a high degree of resolution and vice versa. In other words, the numerical indices reflect the intuitive notion of degree of 'overlap' among different vowel categories in the space defined by the measurements (I.1.3, I.2.3, I.2.5, II.1.2.4, II.2.1.3, II.2.1.5).

After performing the multivariate analyses of variance and the discriminant analyses, a combined ranking for both statistical procedures was made concerning:

- Pisa variety stressed vowel system (I.1.4);
- Pisa variety unstressed vowel system (I.2.4);
- read and semi-spontaneous speech in Pisa stressed vowel system (II.1.1.3, II.1.2.3, II.1.2.5);
- Pisa and Florence stressed vowel system (II.2.1.4, II.2.1.6).

Concerning each of the four conditions, the final rank was calculated by adding the ranking scores of each analysis (i.e. the ranking scores of the Multivariate analyses of variance plus the ranking

scores of the discriminant analyses).[20] Normalizations with high rank scores reduced speaker-specific variance and maintained the vowels' identity.

In order to avoid a proliferation of extensive tables, only the parameter sets obtaining the best and the worst scores and percentages are displayed in the following tables,[21] whereas the intermediate scores and percentages are briefly reported in the footnotes.

As for the graphical representation, the axes in the formant charts have been flipped so that the vowels are arranged as in a traditional tongue-position vowel chart.

Part I

Normalising factors in the Pisa variety (read speech condition)

The analysis of stressed vowel system (I.1) went before the analysis of unstressed vowel system (I.2). In the final part of § I.2 some observations concerning pretonic and postonic vowels were made.

I.1 Stressed vowels

I.1.1 Multivariate analyses of variance

The main rank scores of the MANOVAs are displayed in Table 3.[22]

In Table 3 can be seen that ERB transformations worked well, especially in combination with *f*0-correction. Twenty-seven normalizations ranked below F1 x F2 in Hertz; all the transformations within the Koenig-scale and all the combinations with F3-correction performed poorer than no transformation. The combinations with (F1-*f*0)-correction seemed to perform better than the combinations with (F2-*f*0)-correction and especially with (F2-F1)-correction.

I.1.2 Discriminant analyses

The correct identification main percentages from discriminant analyses (with and without duration) are displayed in Table 4.[23]

Table 3. Multivariate analyses of variance - Pisa read speech sample rank scores.

Scale	Parameter set	Correction	Rank
ERB	F1-*f*0 x F2	+	42
ERB	F1-*f*0 x F2-*f*0	+	41
ERB	F1 x F2	-	40
Log(Hertz)	F1-*f*0 x F2	+	39
Mel	F1-*f*0 x F2	+	38
Mel	F1-*f*0 x F2-*f*0	+	37
Mel	F1 x F2	-	36
Log(Hertz)	F1 x F2	-	35
Log(Hertz)	F1-*f*0 x F2-*f*0	+	34
Bark	F1 x F2	-	33
Bark	F1 x F2-F1	+	32
Bark	F1-*f*0 x F2-*f*0	+	31
Bark	F1-*f*0 x F2	+	30
Hertz	F1-*f*0 x F2	+	29
Hertz	F1 x F2	-	27
Hertz	F1 x F2-F1	+	27
...	...	...	...
Mel	F1-*f*0 x F3-F2	+	3
Koenig-scale	F1-*f*0 x F3-F2	+	2
Hertz	F1-*f*0 x F3-F2	+	1

Table 4. Discriminant analyses - Pisa read speech sample.

Scale	Parameter set	Correction	% Correct identification	+ Duration: % correct identification
Log(Hertz)	F1 x F2-F1	+	94%	94.5%
ERB	F1 x F2-F1	+	94%	94.3%
Log(Hertz)	F1 x F2	-	93.7%	94%
Log(Hertz)	F1-*f*0 x F2	+	93.6%	93.7%
Log(Hertz)	F1-*f*0 x F2-F1	+	93.5%	94%
ERB	F1 x F2	-	93.4%	94%
Mel	F1 x F2-F1	+	93.2%	93.7%
Log(Hertz)	F2-*f*0 x F1-*f*0	+	93.2%	93%
ERB	F1-*f*0 x F2	+	93%	93.5%
ERB	F1-*f*0 x F2-F1	+	92.9%	93.7%
ERB	F2-*f*0 x F1-*f*0	+	92.8%	92.7%
Bark	F1 x F2	-	92.6%	93.3%
Mel	F1 x F2	-	92.6%	93.1%
Bark	F1 x F2-F1	+	92.5%	93.3%
Mel	F1-*f*0 x F2-F1	+	92.4%	92.4%
Mel	F1-*f*0 x F2	+	92.4%	92.5%
Koenig-scale	F1 x F2-F1	+	92.3%	92.7%
Mel	F1-*f*0 x F2-*f*0	+	92.2%	92.3%
Bark	F1-*f*0 x F2	+	92.1%	92.2%
Bark	F1-*f*0 x F2-F1	+	92%	92.3%
Bark	F1-*f*0 x F2-*f*0	+	91.9%	91.9%
Hertz	F1 x F2	-	91.6%	92%
Hertz	F1 x F2-F1	+	91.6%	92%
...	...	...	...	...
Mel	F2-F1 x F3-F2	+	74.3%	76.5%
ERB	F2-F1 x F3-F2	+	72.5%	74.5%
Log(Hertz)	F2-F1 x F3-F2	+	71.6%	73%

The best recognition rates were obtained with the Log(Hz) and the ERB transform; the higher scores were always gained without F3-correction. If duration was also involved in the discrimination tasks, almost all the recognition rates got slightly better.[24]

I.1.3 Timbre confusion

If the Hertz scale is considered, the low and mid-low vowels /ɛ a ɔ/ have the higher recognition rates with the following parameters: F1 x F2, F1 x F2-F1, F1-*f*0 x F2, F1-*f*0 x F2-F1 (see Table 5). The front mid-low vowel /ɛ/ always has the highest recognition rate: this is not odd, considering its status of *shibboleth* in this particular variety of Italian. In the high point vowels, the confusion is only between two categories (e.g. /i/ and /e/; /u/ and /o/). To the exclusion of /a/, the confusion is likely to involve different vowel of the same series (front or back).

In a general sense, front vowels are better recognized than back vowels.

Table 5. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1 x F2, Hz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**87**	**13**	0	0	0	0	0	
/e/	**22.1**	**73.1**	**4.8**	0	0	0	0	
/ɛ/	0	0	**99**	**1**	0	0	0	
/a/	0	0	**0.5**	**96.8**	**2.8**	0	0	
/ɔ/	0	0	0	**4.7**	**89.9**	**5.3**	0	
/o/	0	0	0	0	**4.5**	**86.5**	**9**	
/u/	0	0	0	0	0	**17.4**	**82.6**	
total								**91.6%**

If duration is considered, the discrimination of back vowels (especially /u/) gets better.[25]

A conspicuous change is observable with the parameters F1-*f*0 x F3-F2: with *f*0- and F3-correction, the 'point' vowels /i u a/ have the higher recognition rates. The confusion involves also different vowels of different series (e.g. /ɔ/ could be interpreted as /ɛ/, /o/ could be interpreted as /e/).

Table 6. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1-*f*0 x F3-F2, Hz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**95.9**	**1.6**	0	0	0	0	**2.4**	
/e/	**16.3**	**78.8**	**4.8**	0	0	0	0	
/ɛ/	0	**2.9**	**88**	**9.1**	0	0	0	
/a/	0	0	**5.5**	**90.1**	**4.4**	0	0	
/ɔ/	0	0	**3.6**	**10.7**	**81**	**4.2**	**0.6**	
/o/	0	0	**1.1**	0	**4.5**	**87.6**	**6.7**	
/u/	0	0	0	0	0	**8.7**	**91.3**	
total								**88%**

The recognition task gets worse with F1-F2 x F3-F2, especially for the back vowels (see Table 7). It is worthwhile to underline that even in this framework the vowel /ɛ/ achieves the highest recognition rates.

Table 7. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2, Hz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**83.7**	**16.3**	0	0	0	0	0	
/e/	**24**	**71.2**	**4.8**	0	0	0	0	
/ɛ/	0	**1**	**99**	0	0	0	0	
/a/	0	0	**0.7**	**91.5**	**5.5**	**2.3**	0	
/ɔ/	0	0	0	**37.5**	**56**	**1.2**	**5.4**	
/o/	0	0	**1.1**	**29.2**	**47.2**	**9**	**14.6**	
/u/	0	0	0	**15.9**	**44.9**	**7.2**	**31.9**	
total								**75.7%**

If duration is considered, the discrimination of high and mid-high vowels (especially /u/) gets better:

Table 8. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2 x Duration, Hz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**0.2**	**13.8**	0	0	0	0	0	
/e/	**23.1**	**72.1**	**4.8**	0	0	0	0	
/ɛ/	0	**1**	**98.6**	**0.5**	0	0	0	
/a/	0	0	**0.5**	**91.5**	**5.7**	**2.1**	**0.2**	
/ɔ/	0	0	0	**36.9**	**57.1**	**0.6**	**5.4**	
/o/	0	0	0	**28.1**	**48.3**	**11.2**	**12.4**	
/u/	0	0	0	**15.9**	**36.2**	**1.4**	**46.4**	
total								**77.1%**

As for the auditory scales, the best and the worst classification rates on the Mel scale are displayed in Tables 10-12. The highest recognition rates is gained with the parameters F1 x F2-F1:[26]

Table 9. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1 x F2-F1, Mel values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**98.4**	**1.6**	0	0	0	0	0	
/e/	**15.4**	**79.8**	**4.8**	0	0	0	0	
/ɛ/	0	**2.4**	**97.6**	0	0	0	0	
/a/	0	0	**0.9**	**96.6**	**2.5**	0	0	
/ɔ/	0	0	0	**4.1**	**91.1**	**4.7**	0	
/o/	0	0	0	0	**4.5**	**86.5**	**9**	
/u/	0	0	0	0	0	**17.4**	**82.6**	
total								**93.2%**

If duration is considered, the recognition rates get slightly worse (93.1% total correct recognition), and only the recognition of the vowel /u/ improves (84.1% instead of 82.6%).

The worst recognition rates are achieved with the parameters F2-F1 x F3-F2, especially for the back vowels:

Table 10. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2, Mel values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**0.6**	**15.4**	0	0	0	0	0	
/e/	**25**	**70.2**	**4.8**	0	0	0	0	
/ɛ/	0	**1**	**99**	0	0	0	0	
/a/	0	0	**0.9**	**89.4**	**7.8**	**1.8**	0	
/ɔ/	0	0	0	**36.3**	**60.1**	**3.6**	0	
/o/	0	0	0	**31.5**	**49.4**	**10.1**	**9**	
/u/	0	0	0	**20.3**	**56.5**	**14.5**	**8.7**	
total								**74.3%**

If duration is considered, the recognition rates get slightly better,[27] especially for /i/ (90.2% instead of 84.6%), /e/ (71.2% instead of 70.2%), /a/ (90.1% instead of 89.4%), /o/ (16.9% instead of 10.1%), /u/ (33.3% instead of 8.7%).

A similar trend can be found in the unstressed system (see I.2.3).

I.1.4 Pisa read speech sample composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses are displayed in Table 11.[28]

Table 11. Pisa read speech sample composite rank scores.

Scale	Parameter set	crs
Log(Hertz)	F1-*f*0 x F2	42
ERB	F1 x F2	41
ERB	F1-*f*0 x F2	40
Log(Hertz)	F1 x F2	39
ERB	F1-*f*0 x F2-*f*0	38
Log(Hertz)	F1-*f*0 x F2-*f*0	37
Mel	F1 x F2	36
Mel	F1-*f*0 x F2	35
Bark	F1 x F2	34
Mel	F1-*f*0 x F2-*f*0	33
Bark	F1 x F2-F1	32
Mel	F1 x F2-F1	30
ERB	F1 x F2-F1	30
Bark	F1-*f*0 x F2	29

Scale	Parameter set	crs
Bark	F1-*f*0 x F2-*f*0	27
Log(Hertz)	F1 x F2-F1	27
Log(Hertz)	F1-*f*0 x F2-F1	25
Koenig-scale	F1 x F2-F1	25
Hertz	F1 x F2	20
Hertz	F1 x F2-F1	20
...	...	...
ERB	F2-F1 x F3-F2	3
Hertz	F1-*f*0 x F3-F2	2
Log(Hertz)	F2-F1 x F3-F2	1

Nineteen normalization methods performed poorer than no normalization at all. Log(Hz), ERB, Mel, and Bark combinations performed better than Hz and Koenig-scale combinations. *F*0-correction performed much better than F3-correction.

In Figure 1 the formant data according to one of the best normalization method from Table 11 are displayed.

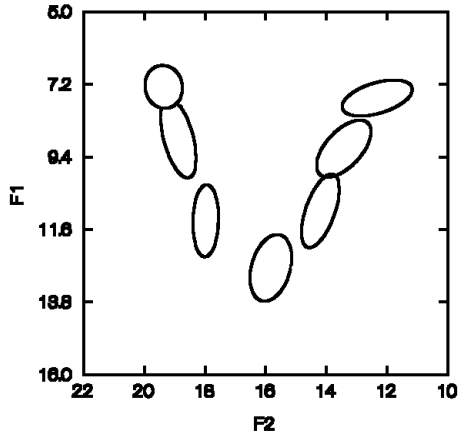

Fig. 1. Regions (68% bivariate ellipsoids) of the Pisa speech sample: F1 x F2, ERB scale.

In Figure 2 the formant data following to one of the worst normalization method from Table 11 are displayed.

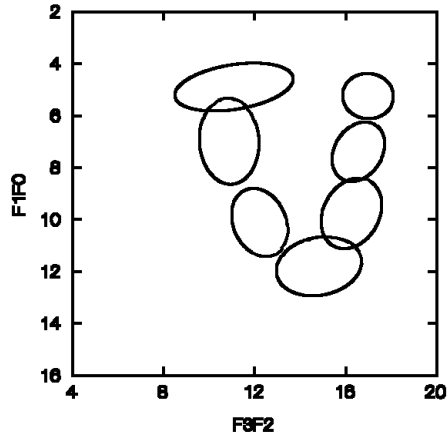

Fig. 2. Regions (68% bivariate ellipsoids) of the Pisa speech sample: F1-*f*0 x F3-F2, ERB scale.

Third formant-correction and *f*0-correction produce more overlapping formant distributions, and larger ellipsoids.

I.2 Unstressed vowels

I.2.1 Multivariate analyses of variance

The main rank scores of the MANOVAs are displayed in Table 12.[29] In comparison to stressed vowels, Hotelling's F traces in the unstressed sample (see Table 2 in the Appendices) are always lower because of the high speaker-individual variability. Generally speaking, standard deviations of the unstressed vowels are (much) higher than standard deviations of stressed vowels.

Table 12. Multivariate analyses of variance - Pisa read speech sample rank scores.

Scale	Parameter set	Correction	Rank
Hertz	F1 x F2	-	41
Hertz	F1 x F2-F1	+	41
...	...	...	...
Log(Hertz)	F1-*f*0 x F3-F2	+	3
Koenig-scale	F1-*f*0 x F3-F2	+	2
Log(Hertz)	F2-F1 x F3-F2	+	1

All the normalization methods performed poorer than no normalization at all. Methods in the Hz scale seemed to work better than non-linear auditory transforms in reducing speaker-specific variance.

I.2.2 Discriminant analyses

The correct identification main percentages from discriminant analyses (with and without duration) are displayed in Table 13:[30]

Table 13. Discriminant analyses - Pisa read speech sample.

Scale	Parameter set	Correction	% Correct identification	+ D: % correct identification
Mel	F1 x F2-F1	+	89.8%	90%
ERB	F1 x F2-F1	+	89.7%	89.9%
Log(Hertz)	F1 x F2-F1	+	89.6%	89.6%
Log(Hertz)	F1 x F2	-	89.5%	89.6%
ERB	F1 x F2	-	89.5%	89.6%
Koenig-scale	F1 x F2-F1	+	89.4%	89.7%
Mel	F1-*f*0 x F2	+	89.3%	89.9%
Koenig-scale	F1 x F2	-	89.2%	88.9%
Bark	F1 x F2	-	89.1%	89.2%
Bark	F1 x F2-F1	+	89.1%	89.1%
Mel	F1-*f*0 x F2-*f*0	+	89%	89.8%
Hertz	F1 x F2	-	89%	89.4%
Hertz	F1 x F2-F1	+	89%	89.4%
...	...	...	...	...
Koenig-scale	F3-F2 x F2-F1	+	69.9%	70.2%
ERB	F2-F1 x F3-F2	+	68.9%	69.5%
Log(Hertz)	F2-F1 x F3-F2	+	68.1%	68.3%

The unstressed vowel system has in the complex lower combined correct recognition rates, compared with the ones from the stressed vowel system; and the confusion involves even all the five vowel categories at a time: it is known that unstressed vowels occupy slightly less peripheral positions in the vowel space. If also duration is involved in the discrimination tasks, almost all the recognition rates get slightly better.[31]

I.2.3 Timbre confusion

The best and the worst performance on the logarithmic scale are displayed in Tables 14-17. On the F1 x F2 scale, the highest score is achieved by /a/.

Table 14. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1 x F2, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**91.1**	**8.3**	**0.6**	0	0	
/e/	**16.1**	**79.6**	**3.9**	0	**0.4**	
/a/	0	**0.4**	**96.6**	**2.5**	**0.4**	
/o/	0	**1.1**	**2.6**	**91.5**	**4.9**	
/u/	0	**3.1**	0	**25.8**	**71.1**	
total						**89.5%**

On the F2-F1 x F3-F2 scale, the high vowels get worse (in particular, /u/ has no classification score), whereas the front mid vowel gets better:

Table 15. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**50.8**	**48.6**	**0.6**	0	0	
/e/	**7.8**	**91.9**	0	**0.4**	0	
/a/	0	**1.9**	**66.1**	**32**	0	
/o/	0	**0.4**	**23.2**	**76.3**	0	
/u/	0	**2.1**	**7.3**	**90.6**	0	
total						**68.1%**

If duration is considered, /e/ and /u/ are slightly better distinguished on the F1 x F2 scale, but the whole recognition rate does not improve.

Table 16. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1 x F2 x D, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**91.1**	**8.3**	**0.6**	0	0	
/e/	**16.5**	**80.7**	**2.5**	0	**0.4**	
/a/	0	**0.4**	**96.6**	**2.5**	**0.4**	
/o/	0	**1.5**	**2.6**	**90.2**	**5.7**	
/u/	0	**2.1**	0	**23.7**	**74.2**	
total						**89.5%**

On the F2-F1 x F3-F2 scale, duration improves /i/ and /u/ distinction, and the whole recognition rate gets slightly better:

Table 17. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2 x D, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**61.5**	**38**	**0.6**	0	0	
/e/	**9.9**	**89.8**	0	**0.4**	0	
/a/	0	**1.7**	**64.8**	**33.3**	**0.2**	
/o/	0	**0.2**	**23.7**	**75.3**	**0.9**	
/u/	0	**2.1**	**5.2**	**91.7**	**1**	
total						**68.3%**

A similar tendency has been noticed in the stressed system (I.1.3).

I.2.4 Pisa read speech sample composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses are displayed in Table 18.[32]

Table 18. Pisa read speech sample composite rank scores.

Scale	Parameter set	crs
Koenig-scale	F1 x F2-F1	42
Mel	F1 x F2-F1	41
Bark	F1 x F2	35
Hertz	F1 x F2	35
Hertz	F1 x F2-F1	35
...	...	...
Koenig-scale	F1-*f*0 x F3-F2	3
ERB	F2-F1 x F3-F2	2
Log(Hertz)	F2-F1 x F3-F2	1

Thirty-four normalization methods performed poorer than no normalization at all. F2-correction seemed to work better than *f*0-correction. F1-*f*0 combination seemed to work better with Mel than with Bark or ERB scales; F2-F1 combination seemed to work better with Koenig, Mel and Hz scales rather than with Bark scale. All the combinations with Log(Hz) ranked lower than F1 x F2 in Hz. Finally,

equivalent rank scores were gained by the following parameter sets: F1 x F2 in Bark, Hertz, Koenig-scale, and ERB; F1 x F2-F1 in Hertz; and F1-*f*0 x F2 in Mel.

In Figures 3 and 4 formant data on the Koenig-scale are displayed.

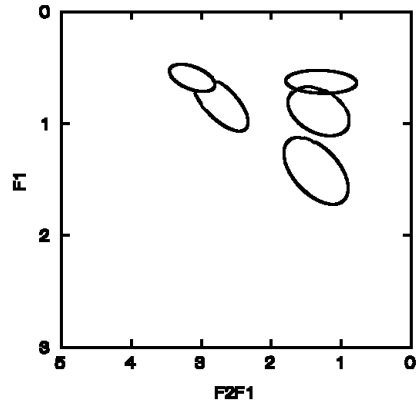

Fig. 3. Regions (68% bivariate ellipsoids) of the Pisa speech sample: F1 x F2-F1, Koenig-scale.

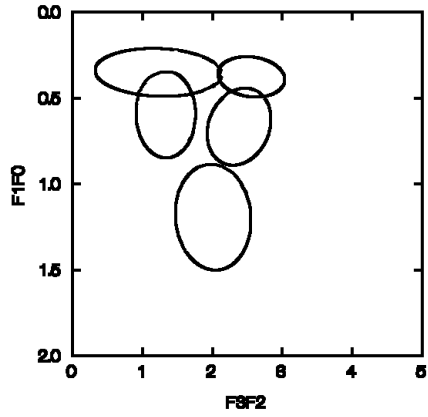

Fig. 4. Regions (68% bivariate ellipsoids) of the Pisa speech sample: F1-*f*0 x F3-F2, Koenig-scale.

F2-correction seems to work better than *f*0- and F3-correction, which produce more overlapping formant distributions, and larger ellipsoids.

I.2.5 Pretonic and postonic vowels. Timbre confusion

The pretonic vowels were divided from the postonic ones, in order to see which vowel categories were better distinguished depending on the stress position (see Tables 3-4 in the Appendices). In the postonic position, high vowels usually obtain the worse recognition rates, as shown in tables 19-22: /i/ falls from 93.5% to 76.7%, /u/ falls from 97% to 43.3%, as far as ERB values are concerned (F1 x F2-F3); /i/ falls from 92.4% to 84.7%, /u/ falls from 95.5% to 32.1%, as far as Mel values are concerned (F1-*f*0 x F2-F1).

Table 19. Confusion matrix from the simulated vowel recognition task in the Pisa read speech - Pretonic Vowels (F1 x F2-F1, ERB values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**93.5**	**5.6**	**0.9**	0	0	
/e/	**13.3**	**85.7**	0	0	**1**	
/a/	0	**2.8**	**96.1**	**1.1**	0	
/o/	**1.8**	**3.5**	**5.3**	**84.2**	**5.3**	
/u/	**1.5**	0	0	**1.5**	**97**	
total						**92.3%**

Table 20. Confusion matrix from the simulated vowel recognition task in the Pisa read speech - Postonic Vowels (F1 x F2-F1, ERB values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**76.7**	**23.3**	0	0	0	
/e/	**15.6**	**83.3**	**1.1**	0	0	
/a/	0	**0.7**	**95.6**	**3.4**	**0.3**	
/o/	0	**1.2**	**1.9**	**94.2**	**2.7**	
/u/	0	**6.7**	0	**50**	**43.3**	
total						**89.8%**

Table 21. Confusion matrix from the simulated vowel recognition task in the Pisa read speech - Pretonic Vowels (F1-*f*0 x F2-F1, Mel values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**92.4**	**6.7**	**1**	0	0	
/e/	**15.3**	**83.7**	0	0	**1**	
/a/	0	**2.8**	**96.1**	**1.1**	0	
/o/	**1.8**	**3.6**	**1.8**	**82.1**	**10.7**	
/u/	0	**1.5**	0	**4.6**	**93.8**	
total						**91%**

Table 22. Confusion matrix from the simulated vowel recognition task in the Pisa read speech - Postonic Vowels (F1-*f*0 x F2-F1, Mel values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/a/	/o/	/u/	total
/i/	**84.7**	**15.3**	0	0	0	
/e/	**15.9**	**83.5**	**0.6**	0	0	
/a/	0	**0.7**	**94.6**	**4.7**	0	
/o/	0	**0.5**	**1.5**	**96.7**	**1.3**	
/u/	0	**3.6**	0	**64.3**	**32.1**	
total						**90.8%**

Almost all the postonic vowels are final ones: the discriminant analyses and confusion matrices confirm a well known tendency concerning unstressed final vowels, which although of long duration due to final lengthening do not move towards target but instead towards a 'schwa' vowel, i.e. coarticulate with a relaxed position of the vocal tract.

Part II

Stylistic and regional variation

II.1 Normalising factors in two different speech styles (Pisa sample)

In order to compare two different speech styles, a sub-*corpus* of the Pisa sample stressed vowel system was compared with a sample of semi-spontaneous speech uttered by the same speakers.

II.1.1 Read speech sample

II.1.1.1 Multivariate analyses of variance

The main rank scores of the MANOVAs are displayed in Table 23.[33]

Table 23. Multivariate analyses of variance - Pisa read speech sample rank scores.

Scale	Parameter set	Correction	Rank
ERB	F1 x F2	-	42
Log(Hertz)	F1 x F2	-	41
Mel	F1 x F2	-	40
Bark	F1 x F2	-	39
Hertz	F1 x F2	-	37
Hertz	F1 x F2-F1	+	37
...	...	...	...
Koenig-scale	F1-*f*0 x F3-F2	+	3
Hertz	F1-*f*0 x F3-F2	+	2
Log(Hertz)	F2-F1 x F3-F2	+	1

Results displayed in Table 23 seem to be somehow different from what has been noticed in I.1.1, although some similar trends are still easily distinguishable (e.g. the validity of the ERB scale). In Table 23 can be seen that ERB, Log(Hz), Mel, and Bark scale work well, especially without *f*0- and formant correction. Thirty-six normalizations ranked below F1 x F2 in Hertz. All the transformations with *f*0-correction and, most of all, with F3-correction performed poorer than no transformation.

II.1.1.2 Discriminant analyses

The correct identification main percentages from discriminant analyses (with and without duration) are displayed in Table 24.[34] The ranking of the different parameter sets is fairly similar to that displayed in Table 4.

Table 24. Discriminant analyses - Pisa read speech sample.

Scale	Parameter set	Correction	% Correct identification	+ D: % correct identification
Log(Hertz)	F1 x F2-F1	+	98.4%	98.1%
ERB	F1 x F2-F1	+	98.1%	98.1%
Log(Hertz)	F1 x F2	-	97.8%	98.4%
Mel	F1 x F2	-	97.5%	98.4%
Mel	F1 x F2-F1	+	97.5%	98.4%
Bark	F1 x F2	-	97.5%	98.1%
ERB	F1 x F2	-	97.5%	98.4%
Bark	F1 x F2-F1	+	97.5%	97.8%
Koenig-scale	F1 x F2-F1	+	97.5%	98.4%
Hertz	F1 x F2	-	97.1%	98.4%
Hertz	F1 x F2-F1	+	97.1%	98.4%
...	...	...	...	...
Mel	F2-F1 x F3-F2	+	73.9%	77.1%
Koenig-scale	F2-F1 x F3-F2	+	73.9%	77.1%
Log(Hertz)	F2-F1 x F3-F2	+	73.9%	73.9%

The best recognition rates are obtained with the Log(Hertz) and the ERB transform; the higher scores are always gained without F3-correction. If duration is also involved in the discrimination tasks, almost all the recognition rates get slightly better.[35]

II.1.1.3 Composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses are displayed in Table 25.[36]

Table 25. Pisa read speech sample composite rank scores.

Scale	Parameter set	crs
Log(Hertz)	F1 x F2	42
ERB	F1 x F2	41
Mel	F1 x F2	40
Bark	F1 x F2	39
Koenig-scale	F1 x F2-F1	38
Hertz	F1 x F2	36
Hertz	F1 x F2-F1	36
...	...	...
Koenig-scale	F1-*f*0 x F3-F2	3
ERB	F2-F1 x F3-F2	2
Log(Hertz)	F2-F1 x F3-F2	1

In comparison with the results displayed in Table 11, which refer to six speakers, some differences and some common trends can

be found. As for the former, *f*0-correction performed poorer in two speakers than in six speakers; as for the latter, in both cases F1 x F2 on the ERB scale reached the same rank score, and F2-F1 x F3-F2 on the Log(Hz) scale reached the lowest rank score.

II.1.2 Semi-spontaneous speech

II.1.2.1 Multivariate analyses of variance

The main rank scores of the MANOVAs are displayed in Table 26.[37] Although F values of semi-spontaneous speech are very different than the F values of read speech (the latter are much higher than the former because of the higher variability of the connected speech tokens), in both styles normalizations rank in a quite similar way.

Table 26. Multivariate analyses of variance - Pisa semi-spontaneous speech sample rank scores.

Scale	Factors	Correction	Rank
ERB	F1 x F2	-	42
Mel	F1 x F2	-	41
Bark	F1 x F2	-	40
Log(Hertz)	F1 x F2	-	39
Hertz	F1 x F2	-	37
Hertz	F1 x F2-F1	+	37
...	...	...	...
Mel	F1-*f*0 x F3-F2	+	3
Hertz	F1-*f*0 x F3-F2	+	2
Koenig-scale	F1-*f*0 x F3-F2	+	1

Thirty-five normalizations ranked below F1 x F2 in Hertz; all the transformations with *f*0- and, most of all, with F3-correction performed poorer than no transformation at all. F2-correction performed better than *f*0-correction.

II.1.2.2 Discriminant analyses

The correct identification main percentages from discriminant analyses (with and without duration) are displayed in Table 27.[38] Vowels are of shorter duration and there is less acoustic vowel contrast in 'informal', conversational speech than in the formal reading

of a word list; vowels are longest and show the greatest acoustic contrast when produced in isolation (see the graphical representations in Figures 5-8).[39] The more overlapping formant distributions cause lowered recognition rates in spontaneous speech. Errors are much more numerous with the semi-spontaneous samples than with the read ones: the combined correct recognition rates reach 98.4% in read speech and only 79% in connected speech.

Table 27. Discriminant analyses - Pisa semi-spontaneous speech sample.

Scale	Parameter set	Correction	% Correct identification	+ D: % correct identification
Log(Hertz)	F1 x F2	-	79%	79.3%
Mel	F1 x F2	-	78.7%	79.8%
Log(Hertz)	F1 x F2-F1	+	78.2%	77.3%
ERB	F1 x F2	-	78.2%	79%
ERB	F1-*f*0 x F2-F1	+	77.9%	79.2%
Mel	F1 x F2-F1	+	77.9%	79.3%
Bark	F1 x F2	-	77.9%	79.6%
ERB	F1 x F2-F1	+	77.6%	76.8%
Hertz	F1-*f*0 x F2-*f*0	+	77.3%	77.3%
Log(Hertz)	F1-*f*0 x F2-*f*0	+	77.3%	78.9%
Hertz	F1-*f*0 x F2-F1	+	77.3%	77.3%
ERB	F1-*f*0 x F2-*f*0	+	77.3%	77.9%
Log(Hertz)	F1-*f*0 x F2-F1	+	77.3%	77.9%
Bark	F1 x F2-F1	+	77.3%	77.9%
Koenig-scale	F1 x F2-F1	+	77.3%	79.3%
Mel	F1-*f*0 x F2-*f*0	+	77.3%	78.5%
Koenig-scale	F1 x F2	-	77.3%	78.7%
Hertz	F1 x F2	-	77.1%	77.9%
Hertz	F1 x F2-F1	+	77.1%	77.9%
...	...	...	...	...
Bark	F2-F1 x F3-F2	+	67.3%	68.2%
ERB	F2-F1 x F3-F2	+	65.2%	66.4%
Log(Hertz)	F2-F1 x F3-F2	+	63.3%	66.1%

The best recognition rates are obtained with F1 x F2 in Log(Hz) and in Mel; in both styles the worst recognition rates is obtained with F2-F1 x F3-F2 in Log(Hz). If the percentages displayed in Table 27 are compared with those of Table 24, it can be seen that duration seems to be slightly more useful in semi-spontaneous speech condition than in read speech condition.[40] Because of centralization and increasing intra-cluster variability, the spectral contrasts are somewhat reduced in semi-spontaneous speech: duration could therefore be more helpful in the classification of vowel categories.

II.1.2.3 Composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses are displayed in Table 28.[41]

Table 28. Pisa semi-spontaneous speech sample composite rank scores.

Scale	Parameter set	crs
Mel	F1 x F2	42
ERB	F1 x F2	40
Log(Hertz)	F1 x F2	40
Bark	F1 x F2	39
Mel	F1 x F2-F1	38
Koenig-scale	F1 x F2	37
Hertz	F1 x F2	34
Hertz	F1 x F2-F1	34
...	...	...
ERB	F2-F1 x F3-F2	3
Log(Hertz)	F2-F1 x F3-F2	1
Koenig-scale	F1-*f*0 x F3-F2	1

Thirty-three normalizations ranked below F1 x F2 in Hz. Without any *f*0- or formant correction, Mel, ERB, Log(Hz), Bark and Koenig-scale performed better than the Hz scale. In semi-spontaneous speech condition, F1 x F2 in Hz performed poorer than in the read speech condition.

In both speech styles, five combinations ranked in the same way,[42] and two couples of combinations behaved alike (F1 x F2 and F1 x F2-F1 in Hz; F1-*f*0 x F2-*f*0 and F1-*f*0 x F2-F1 in Hz). On the whole, if a comparison is made between both styles, a good convergence can be found (the same cannot be said as for the geographical variation: see II.2).

II.1.2.4 Timbre confusion

One of the best combination in both styles is F1 x F2 in Log(Hz):

Table 29. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1 x F2, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**100**	0	0	0	0	0	0	
/e/	**10.7**	**89.3**	0	0	0	0	0	
/ɛ/	0	0	**100**	0	0	0	0	
/a/	0	0	0	**99.1**	**0.9**	0	0	
/ɔ/	0	0	0	**2.1**	**97.9**	0	0	
/o/	0	0	0	0	0	**94.7**	**5.3**	
/u/	0	0	0	0	0	**5**	**95**	
total								**97.8%**

Table 30. Confusion matrix from the simulated vowel recognition task in the Pisa semi-spontaneous speech (F1 x F2, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**88.9**	**9.5**	**1.6**	0	0	0	0	
/e/	**31.4**	**60**	**8.6**	0	0	0	0	
/ɛ/	0	**3.9**	**92.1**	**3.9**	0	0	0	
/a/	0	0	**8.7**	**85.9**	**5.4**	0	0	
/ɔ/	0	0	**2.5**	**20**	**57.5**	**20**	0	
/o/	**3.2**	0	**6.5**	0	**9.7**	**61.3**	**19.4**	
/u/	0	0	0	0	**4**	**24**	**72**	
total								**79%**

The worst combination in both styles is F2-F1 x F3-F2 in Log(Hz):

Table 31. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**69.2**	**30.8**	0	0	0	0	0	
/e/	**7.1**	**92.9**	0	0	0	0	0	
/ɛ/	0	0	**100**	0	0	0	0	
/a/	0	0	**0.9**	**84.8**	**14.3**	0	0	
/ɔ/	0	0	0	**33.3**	**66.7**	0	0	
/o/	0	0	0	**78.9**	**21.1**	0	0	
/u/	0	0	0	**55**	**45**	0	0	
total								**73.9%**

Table 32. Confusion matrix from the simulated vowel recognition task in the Pisa semi-spontaneous speech (F2-F1 x F3-F2, LogHz values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**89.7**	**6.9**	**3.4**	0	0	0	0	
/e/	**44.8**	**31**	**24.1**	0	0	0	0	
/ɛ/	**1.4**	**5.6**	**93**	0	0	0	0	
/a/	0	0	**8.3**	**84.5**	**7.1**	0	0	
/ɔ/	0	0	**2.7**	**67.6**	**29.7**	0	0	
/o/	0	0	**3.7**	**63**	**33.3**	0	0	
/u/	0	0	0	**70.8**	**29.2**	0	0	
total								**63.3%**

In the read speech condition timbre confusion involves two or three categories at time, whereas in the semi-spontaneous speech condition, even five categories at time are involved in the misclassification. It is interesting to notice that even in the connected speech sample the vowel /ɛ/ has always the highest recognition rates. Log(Hz), Mel, ERB, Bark, Koenig-scale (without formant- or *f*0-correction) work well when having to deal with register variation.

II.1.2.5 Pisa read and semi-spontaneous speech samples composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses for both styles (read speech and semi-spontaneous speech) are displayed in Table 33.[43]

Table 33. Pisa read and semi-spontaneous speech samples composite rank scores.

Scale	Parameter set	crs
Log(Hertz)	F1 x F2	41
Mel	F1 x F2	41
ERB	F1 x F2	40
Bark	F1 x F2	39
Koenig-scale	F1 x F2	36
Mel	F1 x F2-F1	36
Koenig-scale	F1 x F2-F1	36
Hertz	F1 x F2	34
Hertz	F1 x F2-F1	34
...	...	...
ERB	F2-F1 x F3-F2	3
Koenig-scale	F1-*f*0 x F3-F2	2
Log(Hertz)	F2-F1 x F3-F2	1

Thirty-three normalizations ranked below F1 x F2 in Hz; formant data in Hz are displayed in Figures 5 (read speech) and 6 (semi-spontaneous speech). As can be seen, between the read and the semi-spontaneous speech condition there are many differences in the dimension and position of the vowel ellipsoids.

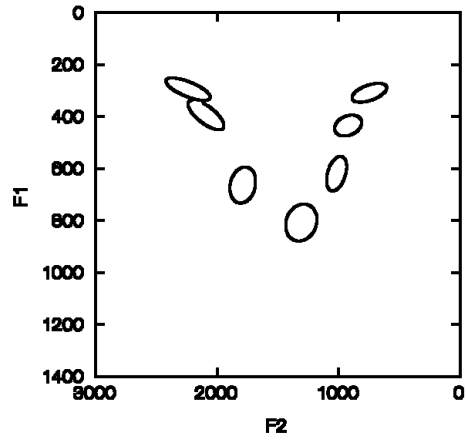

Fig. 5. Regions (68% bivariate ellipsoids) of read speech vowels: F1 x F2, Hz scale.

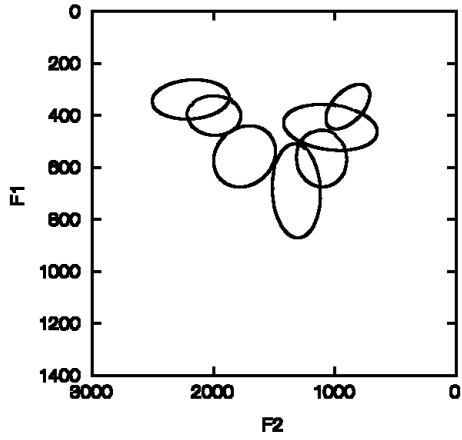

Fig. 6. Regions (68% bivariate ellipsoids) of semi-spontaneous speech vowels: F1 x F2, Hz scale.

All the transformations with *f*0- and, most of all, with F3-correction performed poorer than no transformation at all. Highest ranks

are obtained without formant correction: F1 x F2 in Log(Hz), Mel, ERB, and Bark performed best.

In Figures 7-8 the formant data according to the best normalization method from Table 33 are displayed.

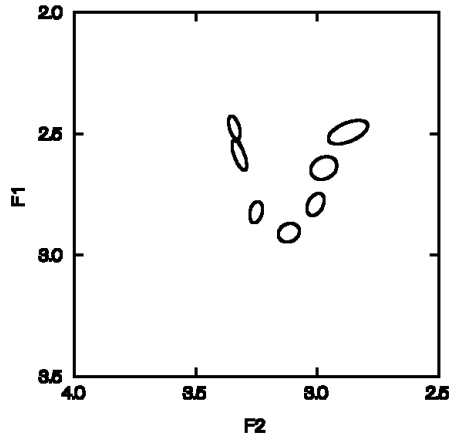

Figure 7. Regions (68% bivariate ellipsoids) of read speech vowels: F1 x F2, Log(Hz) scale.

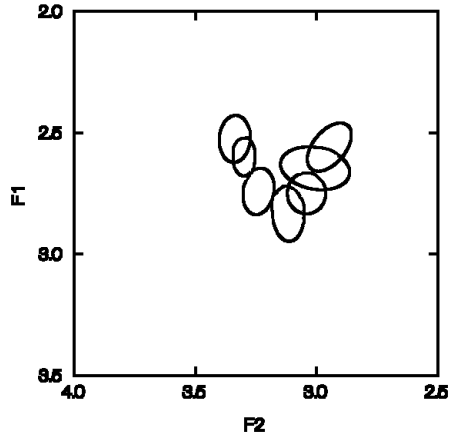

Figure 8. Regions (68% bivariate ellipsoids) of semi-spontaneous speech vowels: F1 x F2, Log(Hz) scale.

With the logarithmic transform, vowel ellipsoids are smaller in both styles.

II.2 Normalizing factors in two different varieties (Pisa and Florence). Read speech samples (sub-corpus)

A great amount of speaker specific variance may obscure the view of different phonetic qualities of the vowels. The main question addressed in this part concerns geographical variation: what aspects of the acoustic measurements reflect genuine phonetic quality differences between the two varieties?

II.2.1 Florence

II.2.1.1 Multivariate analyses of variance

The main rank scores of the MANOVAs are displayed in Table 34.[44] As can be seen in Appendices (Table 7), Florence sample normalizations show in the complex lower F values, probably because of higher standard deviations: the Florence data contain more variability than the Pisa data, possibly indicating lower homogeneity in the Florence speakers.

Table 34. Multivariate analyses of variance - Florence read speech sample rank scores.

Scale	Parameter set	Correction	Rank
Log(Hertz)	F1-*f*0 x F2	+	42
ERB	F1-*f*0 x F2	+	41
Bark	F1-*f*0 x F2	+	40
ERB	F1-*f*0 x F2-*f*0	+	39
Mel	F1-*f*0 x F2	+	38
Bark	F1-*f*0 x F2-*f*0	+	37
Mel	F1-*f*0 x F2-*f*0	+	36
Koenig-scale	F1-*f*0 x F2-*f*0	+	35
Log(Hertz)	F1-*f*0 x F2-*f*0	+	34
Koenig-scale	F1-*f*0 x F2	+	33
Log(Hertz)	F1 x F2	-	32
ERB	F1 x F2	-	31
Bark	F1 x F2	-	30
Mel	F1 x F2	-	29
Koenig-scale	F1 x F2	-	28
Hertz	F1-*f*0 x F2	+	27
Hertz	F1-*f*0 x F2-*f*0	+	25
Hertz	F1-*f*0 x F2-F1	+	25
Hertz	F1 x F2	-	23
Hertz	F1 x F2-F1	+	23
...	...	...	...
Bark	F2-F1 x F3-F2	+	3
ERB	F2-F1 x F3-F2	+	2
Log(Hertz)	F2-F1 x F3-F2	+	1

Twenty-two normalizations ranked below F1 x F2 in Hz, which is not as good as in the Pisa sample in reducing variability. *F*0-correction works very well on every scale in reducing speaker-specific variance.

II.2.1.2 Discriminant analyses

The correct identification main percentages from discriminant analyses (with and without duration) are displayed in Table 35:[45]

Table 35. Discriminant analyses - Florence read speech sample.

Scale	Parameter set	Correction	% Correct identification	+ D: % correct identification
Mel	F1-*f*0 x F2	+	92.9%	92.6%
Log(Hertz)	F1-*f*0 x F2	+	92.9%	92.6%
Mel	F1-*f*0 x F2-*f*0	+	92.6%	92%
ERB	F1-*f*0 x F2	+	92.3%	92.3%
Koenig-scale	F1-*f*0 x F2	+	92.3%	92.3%
Koenig-scale	F1-*f*0 x F2-*f*0	+	91.6%	92.3%
Hertz	F1-*f*0 x F2	+	91.3%	93.2%
Hertz	F1-*f*0 x F2-F1	+	91.3%	92.3%
ERB	F1-*f*0 x F2-*f*0	+	91.3%	91.6%
Hertz	F1-*f*0 x F2-*f*0	+	91.3%	92.3%
Log(Hertz)	F1-*f*0 x F2-*f*0	+	91.3%	91.6%
Koenig-scale	F1-*f*0 x F2-F1	+	91%	90.4%
Mel	F1-*f*0 x F2-F1	+	90.7%	90.4%
ERB	F1 x F2	-	90.4%	90.7%
Hertz	F1 x F2	-	90.4%	91%
Hertz	F1 x F2-F1	+	90.4%	91%
...	...	...	...	...
ERB	F2-F1 x F3-F2	+	70.8%	72.7%
Log(Hertz)	F2-F1 x F3-F2	+	68.3%	69.9%
Bark	F2-F1 x F3-F2	+	67.7%	70.2%

The best recognition rates are always obtained with *f0*-correction in almost all the scales. As already noticed for the Pisa sample, the worst recognition rates are obtained with F3-correction; the percentages of correct identification with F3-correction get better only if also *f*0-correction is added.

II.2.1.3 Timbre confusion in the Florence sample

A good combination in the Florence sample is (F1-*f*0) x (F2-F1) in Bark:

Table 36. Confusion matrix from the simulated vowel recognition task in the Florence read speech (F1-*f*0 x F2-F1, Bark values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**100**	0	0	0	0	0	0	
/e/	**27.8**	**66.7**	**5.6**	0	0	0	0	
/ɛ/	0	**19.3**	**80.7**	0	0	0	0	
/a/	0	0	**3.5**	**94.7**	**1.8**	0	0	
/ɔ/	0	0	0	0	**81.3**	**18.8**	0	
/o/	0	0	0	0	**18.2**	**72.7**	**9.1**	
/u/	0	0	0	0	0	**9.1**	**90.9**	
total								**85.8%**

If duration is considered, the recognition rates get slightly better,[46] especially for /e/ (75% instead of 66.7%), for /ɛ/ (82.5% instead of 80.7%), for /ɔ/ (83.3% instead of 81.3%).

II.2.1.4 Florence read speech sample composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses for the Florence sample are displayed in Table 37.[47]

Table 37. Florence read speech sample composite rank scores.

Scale	Parameter set	crs
Log(Hertz)	F1-*f*0 x F2	42
ERB	F1-*f*0 x F2	40
Mel	F1-*f*0 x F2	40
Mel	F1-*f*0 x F2-*f*0	39
Koenig-scale	F1-*f*0 x F2-*f*0	38
ERB	F1-*f*0 x F2-*f*0	36
Koenig-scale	F1-*f*0 x F2	36
Log(Hertz)	F1-*f*0 x F2-*f*0	35
Bark	F1-*f*0 x F2	34
Hertz	F1-*f*0 x F2	32
Bark	F1-*f*0 x F2-*f*0	32
ERB	F1 x F2	29
Hertz	F1-*f*0 x F2-*f*0	29
Hertz	F1-*f*0 x F2-F1	29

Scale	Parameter set	crs
Log(Hertz)	F1 x F2	28
Koenig-scale	F1 x F2	27
Koenig-scale	F1-*f*0 x F2-F1	26
Mel	F1 x F2	25
Bark	F1 x F2	23
Mel	F1-*f*0 x F2-F1	23
Hertz	F1 x F2	21
Hertz	F1 x F2-F1	21
...	...	...
ERB	F2-F1 x F3-F2	3
Bark	F2-F1 x F3-F2	2
Log(Hertz)	F2-F1 x F3-F2	1

In comparison with the Pisa sample, F1 x F2 in Hz performed more poorly: only twenty normalizations ranked lower than no normalization at all. *F*0-correction performed better than F1 alone. The Mel and the ERB scales performed better than the Bark scale.

On the whole, no convergence can be found between Florence composite rank scores and Pisa composite rank scores: only one parameter set (the worst in both cases) gained the same score (1).

II.2.1.5 Timbre confusion: Florence and Pisa samples

One of the best combination in both varieties is F1 x F2 on the ERB scale, as can be seen in Tables 38-39:

Table 38. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F1 x F2, ERB values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**100**	0	0	0	0	0	0	
/e/	**14.3**	**85.7**	0	0	0	0	0	
/ɛ/	0	0	**100**	0	0	0	0	
/a/	0	0	0	**99.1**	**0.9**	0	0	
/ɔ/	0	0	0	**2.1**	**97.9**	0	0	
/o/	0	0	0	0	0	**94.7**	**5.3**	
/u/	0	0	0	0	0	**5**	**95**	
total								**97.5%**

Table 39. Confusion matrix from the simulated vowel recognition task in the Florence read speech (F1 x F2, ERB values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**96**	**4**	0	0	0	0	0	
/e/	**5.6**	**83.3**	**11.1**	0	0	0	0	
/ɛ/	0	**7**	**93**	0	0	0	0	
/a/	0	0	**1.8**	**96.5**	**1.8**	0	0	
/ɔ/	0	0	0	0	**89.6**	**10.4**	0	
/o/	0	0	0	0	**36.4**	**59.1**	**4.5**	
/u/	0	0	0	0	0	**9.1**	**90.9**	
total								**90.4%**

Florence mid-high vowels seem to be less differenciated. In the Florence variety /ɛ/ is sometimes computed as /e/; in the Pisa variety /ɛ/ has often the highest recognition rates and in any case would be computed as /a/ rather than /e/, since its phonetic value is often [æ] instead of [ɛ]. The vowels better distinguished are /i/ and /ɛ/ in the

Pisa sample, /a/ and /i/ in the Florence sample. The shifting of /a/ into /ɑ/ in the Pisa variety is indirectly proven by the confusion matrices, where /a/ is confused only with /ɔ/, whereas in the Florence variety it is confused also with /ɛ/.

One of the worst combination in both varieties is F2-F1 x F3-F2 on the ERB scale:

Table 40. Confusion matrix from the simulated vowel recognition task in the Pisa read speech (F2-F1 x F3-F2, ERB values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**76.9**	**23.1**	0	0	0	0	0	
/e/	**7.1**	**92.9**	0	0	0	0	0	
/ɛ/	0	0	**100**	0	0	0	0	
/a/	0	0	**0.9**	**83.9**	**15.2**	0	0	
/ɔ/	0	0	0	**31.3**	**68.8**	0	0	
/o/	0	0	0	**73.7**	**26.3**	0	0	
/u/	0	0	0	**55**	**45**	0	0	
total								**74.5%**

Table 41. Confusion matrix from the simulated vowel recognition task in the Florence read speech (F2-F1 x F3-F2, ERB values). Actual groups are in rows, and predicted group membership in columns.

V	/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/	total
/i/	**44**	**52**	**4**	0	0	0	0	
/e/	**5.6**	**66.7**	**27.8**	0	0	0	0	
/ɛ/	0	**7.1**	**92.9**	0	0	0	0	
/a/	0	0	**2.7**	**84.1**	**13.3**	0	0	
/ɔ/	0	0	0	**14.6**	**83.3**	0	**2.1**	
/o/	0	0	0	**22.7**	**72.7**	0	**4.5**	
/u/	0	0	0	**4.5**	**68.2**	0	**27.3**	
total								**70.8%**

In both varieties /o/ cannot be correctly distinguished; in the Pisa variety /u/ also does not reach any recognition score. What is also relevant to observe is the highest identification score reached by /ɛ/ in the Pisa speech.[48]

II.2.1.6 Florence and Pisa read speech samples composite rank scores

The main composite rank scores (*crs*) of the MANOVAs plus discriminant analyses for both varieties are displayed in Table 42.[49]

Table 42. Florence and Pisa read speech samples composite rank scores.

Scale	Parameter set	crs
Mel	F1-*f*0 x F2-*f*0	42
ERB	F1 x F2	40
ERB	F1-*f*0 x F2-*f*0	40
Log(Hertz)	F1 x F2	39
Mel	F1 x F2	38
Koenig-scale	F1 x F2	37
Bark	F1-*f*0 x F2-*f*0	35
Log(Hertz)	F1-*f*0 x F2-*f*0	35
Mel	F1-*f*0 x F2	32
ERB	F1-*f*0 x F2	32
Bark	F1 x F2	32

Scale	Parameter set	crs
Log(Hertz)	F1-*f*0 x F2	31
Mel	F1 x F2-F1	29
Koenig-scale	F1-*f*0 x F2-*f*0	29
Hertz	F1 x F2	26
Hertz	F1 x F2-F1	26
...	...	...
Hertz	F1-*f*0 x F3-F2	3
ERB	F2-F1 x F3-F2	2
Log(Hertz)	F2-F1 x F3-F2	1

Twenty-seven normalizations ranked below F1 x F2 in Hz, whose formant data are displayed in Figures 9 as for the Florence variety.[50]

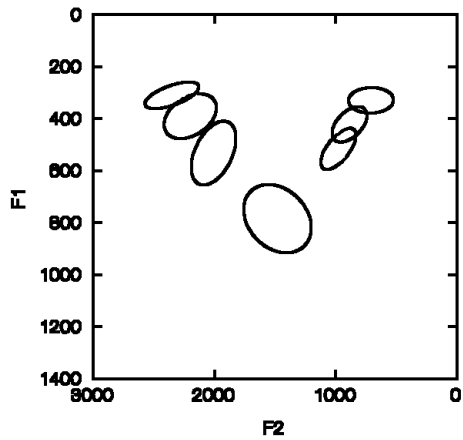

Fig. 9. Regions (68% bivariate ellipsoids) of the Florence vowels: F1 x F2, Hz scale.

Mel, ERB, Log(Hz), and Bark transformations (with and without *f*0-correction) work well when having to deal with data that has relatively little anatomical-physiological variation. As already noted, combinations with F3-correction perform poorly.

Formant data coming from the best normalization are displayed in Figures 10 and 11.

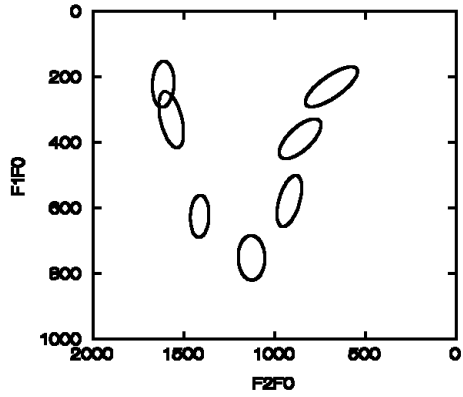

Fig. 10. Regions (68% bivariate ellipsoids) of the Pisa vowels: F1-*f*0 x F2-*f*0, Mel scale.

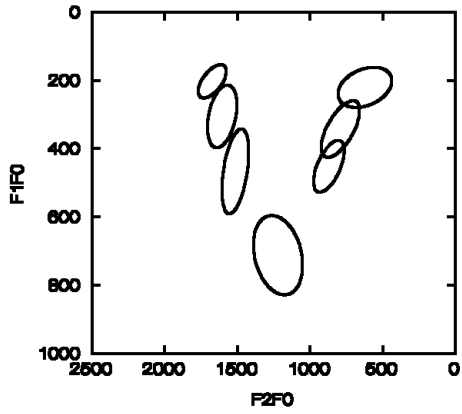

Fig. 11. Regions (68% bivariate ellipsoids) of the Florence vowels: F1-*f*0 x F2-*f*0, Mel scale.

As is shown in Tables 25 and 37, some differences between the Pisa and the Florence samples can be found. In the first one, the best normalization procedures are those without *f*0-correction. The opposite can be observed in the Florence sample. We could argue that the high standard deviations in the Florence sample somehow weaken the normalization procedures,[51] but this assumption would need further research.

III Summary and general discussion

This paper has evaluated the advantages and disadvantages of different potentially normalizing factors in vowel representation. Multivariate analyses of variance were used in order to minimize within-vowel category differences among talkers, whereas discriminant analyses were used in order to maximize vowel-category separability. Composite ranking scores combining MANOVAs and discriminant analyses ranking scores were finally computed: normalization factors with high composite rank scores reduced speaker-specific variance and maintained vowel identity.

Although sex and age variation have not been considered in the present research, dramatic differences in the MANOVAs and in the discriminant rank scores, and in the composite rank scores concerning stress were found: as for the stressed vowels in the Pisa read speech sample, the better results were gained by (F1-*f*0) x F2 on the Log(Hz) scale and by F1 x F2 on the ERB scale; as for the unstressed vowels in the Pisa read speech sample, the better results were gained by F1 x (F2-F1) on the Koenig-scale and by F1 x (F2-F1) on the Mel scale. If geographical variation was considered, *f*0-correction seemed effective in the Florence sample, but ineffective in the Pisa sample. Some discrepancies may have been produced by some kind of weakness in the experimental design (artifacts of measurement, different number of tokens in each vowel category, relatively high values of standard deviations caused by different consonantal context), or by the weakness of the intrinsic normalization procedures tested here, which does not appear capable of reducing variability between speakers to a great degree.

If the statistical results of the whole Pisa sample (both stressed and unstressed vowel systems) are considered, a greater similarity could be found in the scores of the discriminant analyses rather than in the scores of the MANOVAs. In the first case, the highest scores were reached by the following parameter sets, both in the stressed and in the unstressed condition: F1 x (F2-F1) in Log(Hz) and in ERB transform, F1 x F2 in Log(Hz) transform;[52] in the second case, the highest score reached by (F1-*f*0) x F2 in ERB transform in the stressed condition (42) was not confirmed in the unstressed condition, where the best result was surprisingly gained by F1 x F2 in raw Hertz.[53] It follows that intrinsic vowel normalization procedures seemed effective at maximizing differences between vowel categories but seemed somehow deficient in minimization differences in the same vowel spoken by different speakers.[54] Therefore extrinsic nor-

malization procedures should also be tested in the same sample, in order to verify whether they are capable of reducing between-speaker variability to a greater amount.

Some other findings should be noted and summarized as follows.

First of all, across the five non-linear transforms tested, minimal differences were found, and this is perhaps not surprising given the considerable similarity of these transforms.[55]

Second, comparison across different styles within the same speakers seemed to be more successful than comparison across different dialects and different speakers, although, in the second case, an identical type of speech material was used. In other words, trends were more easily picked out in different styles within the same speakers than in different speakers enunciating the same speech material. The same could not be said when stress was concerned: normalization procedures in the same speakers in different stress conditions (i.e. stressed and unstressed) did not reveal the same tendencies.

Third, all the statistical analyses run in the experiment pointed out that formant correction or *f*0-correction were not unconditionally better or worse than the untransformed data, either in Hz or in psychophysical scaling.

In addition, the parameter set F1 x F2 in Hz was exactly equivalent to F1 x (F2-F1) in Hz in every sample and in every comparison (both stylistic and diatopic). Therefore F2-correction did not improve vowel classification when the raw Hz scale was considered.

Another finding concerned the role of the upper formants: F3-correction never improved vowel classification. This was an expected result, since the third formant is not very useful in the Italian vowel systems, which are lacking in rhotacized vowels, front rounded vowels, and nasalized vowels.

As a whole, the parameter set F1 x F2 in ERB seemed to be one of the most successful in reducing inter-speaker variability and in preserving vowel-category separability; the composite rank scores it reached were: 41 (Pisa stressed vowel system in the read speech condition – six speakers), 35 (Pisa unstressed vowel system in the read speech condition – six speakers), 41 (Pisa stressed vowel system in the read speech condition – two speakers), 40 (Pisa stressed vowel system in the semi-spontaneous speech condition – two speakers), 40 (Pisa stressed vowel system in both speech styles), 29 (Florence stressed vowel system – two speakers), 40 (Florence and Pisa stressed vowel system – four speakers). Two other parameter sets

which seemed effective both for stylistic and geographical variation were F1 x F2 on the Log(Hz) and on the Mel scale: with respect to stylistic variation, F1 x F2 in Log(Hz) scored 41 and F1 x F2 in Mel scored 40 (see Table 33); with respect to geographical variation, F1 x F2 in Log(Hz) scored 39 and F1 x F2 in Mel scored 38 (see Table 42).

Finally, with respect to the statistical results in all the different conditions, a greater similarity could be detected in the lowest scores comparing to the highest scores: in every condition the parameter set (F2-F1) x (F3-F2) on the Log(Hz) scale showed the worst performance. Some normalization procedures seemed therefore always weak and ineffective, regardless of speech style and geographical variation.

As for timbre confusion emerging from discriminant analyses, some differences were noted in the four different speech conditions (stressed condition, unstressed condition, speech style, language variety). The unstressed condition had, compared with the stressed condition, overall lower combined correct recognition rates; and the confusion involved even all the five vowel categories at a time, because of a well known centralization tendency occurring in unstressed vowel systems. As for the stressed condition, in the Pisa read speech, the front mid-low vowel /ɛ/ very often got the highest recognition rate in the simulated vowel recognition tasks; and this finding was not confirmed in the Florence read speech, where the vowel /ɛ/ has not a status of *shibboleth*. In the Pisa read speech condition timbre confusion involved two or three categories at time, whereas in the Pisa semi-spontaneous speech condition, even five categories at time were involved in the misclassification, because of centralization and increasing intra-cluster variability; nevertheless, even in the connected speech sample the vowel /ɛ/ got the highest recognition rates.

Besides the normalizations (as discussed above), another source of information was introduced: spectral information versus spectral plus duration information. Some trends were clearly detectable: including duration improved the classification process (in the discriminant analyses the correct identification percentages are slightly higher) but did not reduce inter-speaker variance (Hotelling's F values were considerably lower).[56] The classification results for the spectral and duration information were very often slightly better than those for the spectral information; but the differences were always small, as already noticed in Zahorian & Jagharghi (1993:1977) and in Hillenbrand *et al.* (1995:3109). Durational information seemed not to be enough if the pattern of spectral change throughout the course of

the entire vowel was not considered.[57] The measurements taken in the present samples concerned only the steady state of the vowels[58] and did not include the entire formant trajectories: it is widely known that phonetically relevant information is lost when vowel spectra are reduced to formant representations in the steady state of the segment. The identity of a vowel is determined not only by the formant frequencies at the point of closest approach to target, but also by the direction and rate of adjacent formant transitions.[59] Dynamic properties play an important role in vowel perception: the next step of our research will therefore move into this direction.

Address of the author:

Silvia Calamai, Laboratorio di Linguistica Scuola Normale Superiore, Piazza dei Cavalieri, 7 - 56126 Pisa. E-mail: calsilvia@tiscali.it

Appendices

Normalising factors in the Pisa variety (read speech)

Stressed vowels[60]

Table 1. Multivariate analyses of variance.

Scale	Parameter set	Correction	Hotelling's F	+ Duration: Hotelling's F
ERB	F1-*f*0 x F2	+	3328.347	2253.803
ERB	F1-*f*0 x F2-*f*0	+	3290.335	2235.756
ERB	F1 x F2	-	3283.961	2229.086
Log(Hertz)	F1-*f*0 x F2	+	3259.532	2203.884
Mel	F1-*f*0 x F2	+	3223.062	2184.780
Mel	F1-*f*0 x F2-*f*0	+	3213.147	2186.411
Mel	F1 x F2	-	3210.904	2181.210
Log(Hertz)	F1 x F2	-	3210.770	2177.215
Log(Hertz)	F1-*f*0 x F2-*f*0	+	3210.519	2177.426
Bark	F1 x F2	-	3188.349	2166.129
Bark	F1 x F2-F1	+	3182.423	2162.261
Bark	F1-*f*0 x F2-*f*0	+	3163.279	2148.685
Bark	F1-*f*0 x F2	+	3133.946	2121.503
Hertz	F1-*f*0 x F2	+	3090.553	2092.533
Hertz	F1 x F2	-	3080.983	2089.779
Hertz	F1 x F2-F1	+	3080.983	2089.779
...	...	...	...	...
Mel	F1-*f*0 x F3-F2	+	1170.185	783.061
Koenig-scale	F1-*f*0 x F3-F2	+	1056.639	707.661
Hertz	F1-*f*0 x F3-F2	+	1035.042	693.249

Unstressed vowels[61]

Table 2. Multivariate analyses of variance.

Scale	Parameter set	Correction	Hotelling's F	+ Duration: Hotelling's F
Hertz	F1 x F2	-	1944.943	1316.175
Hertz	F1 x F2-F1	+	1944.943	1316.175
...	...	...	...	...
Log(Hertz)	F1-*f*0 x F3-F2	+	976.038	661.904
Koenig-scale	F1-*f*0 x F3-F2	+	911.080	625.076
Log(Hertz)	F2-F1 x F3-F2	+	894.123	610.052

Unstressed pretonic vowels[62]

Table 3. Discriminant analyses.

Scale	Parameter set	Correction	% Correct identification
Log(Hertz)	F1 x F2	-	92.8%
ERB	F1-*f*0 x F2-*f*0	+	92.4%
ERB	F1 x F2-F1	+	92.3%
Bark	F1-*f*0 x F3-F2	+	92.2%
Log(Hertz)	F1-*f*0 x F2-*f*0	+	92.2%
Log(Hertz)	F1 x F2-F1	+	91.9%
ERB	F1 x F2	-	91.9%
Mel	F1-*f*0 x F2-*f*0	+	91.8%
Log(Hertz)	F1-*f*0 x F2	+	91.6%
ERB	F1-*f*0 x F3-F2	+	91.6%
Bark	F1-*f*0 x F2-*f*0	+	91.6%
Mel	F1 x F2	-	91.5%
ERB	F1-*f*0 x F2-F1	+	91.4%
ERB	F1-*f*0 x F2	+	91.4%
Bark	F1 x F2	-	91.3%
Bark	F1 x F2-F1	+	91.3%
Log(Hertz)	F1-*f*0 x F3-F2	+	91.2%
Mel	F1-*f*0 x F3-F2	+	91.2%
Mel	F1-*f*0 x F2-F1	+	91%
Log(Hertz)	F1-*f*0 x F2-F1	+	90.8%
Mel	F1 x F2-F1	+	90.8%
Mel	F1-*f*0 x F2	+	90.8%
Bark	F1-*f*0 x F2-F1	+	90.6%
Koenig-scale	F1 x F2-F1	+	90.5%
Bark	F1-*f*0 x F2	+	90.4%
Koenig-scale	F2-F1 x F1-*f*0	+	90.4%
Hertz	F1-*f*0 x F3-F2	+	90.2%
Hertz	F1 x F2	-	90.1%
Hertz	F1 x F2-F1	+	90.1%
...	...	...	...
Koenig-scale	F3-F2 x F2-F1	+	72%
ERB	F2-F1 x F3-F2	+	71.4%
Log(Hertz)	F2-F1 x F3-F2	+	68.7%

Unstressed postonic vowels[63]

Table 4. Discriminant analyses.

Scale	Parameter set	Correction	% Correct identification
Koenig-scale	F1 x F2	+	91.5%
Mel	F1-*f*0 x F2	+	91.4%
Hertz	F1 x F2	-	91.2%
Hertz	F1 x F2-F1	+	91.2%
Koenig-scale	F1 x F2-F1	+	91.1%
Mel	F1 x F2	-	90.8%
Mel	F1 x F2-F1	+	90.8%
Mel	F1-*f*0 x F2-F1	+	90.8%
Koenig-scale	F2-F1 x F1-*f*0	+	90.8%
Bark	F1 x F2	-	90.6%
Bark	F1 x F2-F1	+	90.6%
...	...	...	...
Bark	F2-F1 x F3-F2	+	72.8%
Hertz	F2-F1 x F3-F2	+	72.3%
Koenig-scale	F3-F2 x F2-F1	+	72.1%

Pisa samples - stressed vowels (read speech - 2 speakers)[64]

Table 5. Multivariate analyses of variance.

Scale	Parameter set	Correction	Hotelling's F	+ Duration: Hotelling's F
ERB	F1 x F2	-	1426.764	952.642
Log(Hertz)	F1 x F2	-	1421.824	949.484
Mel	F1 x F2	-	1412.439	942.717
Bark	F1 x F2	-	1398.959	933.875
Hertz	F1 x F2	-	1393.297	929.632
Hertz	F1 x F2-F1	+	1393.297	929.632
...	...	...	...	...
Koenig-scale	F1-*f*0 x F3-F2	+	494.054	338.130
Hertz	F1-*f*0 x F3-F2	+	468.541	340.628
Log(Hertz)	F2-F1 x F3-F2	+	413.274	277.254

Pisa samples - stressed vowels (semi-spontaneous speech - 2 speakers)[65]

Table 6. Multivariate analyses of variance.

Scale	Parameter set	Correction	Hotelling's F	+ Duration: Hotelling's F
ERB	F1 x F2	-	294.724	200.928
Mel	F1 x F2	-	293.167	199.319
Bark	F1 x F2	-	292.296	199.093
Log(Hertz)	F1 x F2	-	292.025	198.918
Hertz	F1 x F2	-	282.756	191.312
Hertz	F1 x F2-F1	+	282.756	191.312
...	...	...	...	...
Mel	F1-*f*0 x F3-F2	+	124.672	85.274
Koenig-scale	F1-*f*0 x F3-F2	+	114.511	77.918
Hertz	F1-*f*0 x F3-F2	+	114.909	78.230

Florence samples - stressed vowels (read speech - 2 speakers)[66]

Table 7. Multivariate analyses of variance.

Scale	Parameter set	Correction	Hotelling's F	+ Duration: Hotelling's F
Log(Hertz)	F1-*f*0 x F2	+	591.712	415.052
ERB	F1-*f*0 x F2	+	587.431	415.446
Bark	F1-*f*0 x F2	+	574.074	410.681
ERB	F1-*f*0 x F2-*f*0	+	572.551	401.876
Mel	F1-*f*0 x F2	+	570.300	410.680
Bark	F1-*f*0 x F2-*f*0	+	568.020	405.008
Mel	F1-*f*0 x F2-*f*0	+	565.827	406.646
Koenig-scale	F1-*f*0 x F2-*f*0	+	563.884	409.168
Log(Hertz)	F1-*f*0 x F2-*f*0	+	562.710	390.008
Koenig-scale	F1-*f*0 x F2	+	560.369	404.442
Log(Hertz)	F1 x F2	-	545.376	385.658
ERB	F1 x F2	-	545.022	388.223
Bark	F1 x F2	-	543.269	390.565
Mel	F1 x F2	-	540.481	391.253
Koenig-scale	F1 x F2	-	537.664	389.518
Hertz	F1-*f*0 x F2	+	537.470	393.338
Hertz	F1-*f*0 x F2-*f*0	+	535.444	392.156
Hertz	F1-*f*0 x F2-F1	+	535.444	392.156
Hertz	F1 x F2	-	515.111	378.784
Hertz	F1 x F2-F1	+	515.111	378.784
...	...	...	...	...
Bark	F2-F1 x F3-F2	+	280.728	203.080
ERB	F2-F1 x F3-F2	+	247.670	177.982
Log(Hertz)	F2-F1 x F3-F2	+	222.297	159.045

Notes

* The author would like to thank two anonymous reviewers for their helpful comments on an earlier version of this paper; Elgin K. Eckert for linguistic advice; Maddalena Agonigi and Irene Ricci (Laboratorio di Linguistica, Scuola Normale Superiore) for statistical advice.

1 It is known that back vowels are in better correspondence with the way in which they are heard when one relates backness to the difference between the first and second formant frequencies, rather than to the second formant frequency alone (Lindau 1978).

2 I.e. his physical anatomy, his age and gender, his emotional state.

3 The main models of vowel perception (dynamic-specification versus elaborate target models) are described in Van Son (1993).

4 The classification comes from Ainsworth (1975); see Nearey (1989) for a review. Extrinsic specifications are used in Ladefoged & Broadbent (1957), Gerstman (1968), Lobanov (1971), Nearey (1977), Wakita (1977). The normalization algorithms are sometimes also called 'external' and 'internal': external normalizations involve the attempt to find one or more constants which map one absolute system onto another, internal normalizations carry out a system-dependent, immanent normalization. In this grouping, Nearey's normalization is considered an internal one, although an extrinsic method is used.

5 Some doubts on the use of the fully Log-scale have been raised (Rosner & Pickering 1994:17; Iivonen 1994:75), since the area between 200-500 Hz is enlarged too much compared to the area between 500-800 Hz.

6 The acoustic values have been converted to the Mel scale using the technical approximation from Fant (1973:48).

7 The formula from Traunmüller (1990) was used for the bark-scale transform.

8 The Bark scale assumes a rectangular filter shape whereas the ERB-rate scale "adopts the Roex filter shape derived by Patterson *et al*. (1982) from masking data" (Rosner & Pickering 1994:18). The ERB-rate scale is sometimes preferred to the Bark scale: "in particular, Patterson et al. made no assumptions about filter shape prior to their experiments, whereas the bark scale assumes a rectangular filter shape from the start" (Rosner & Pickering 1994:19).

9 Potter & Steinberg (1950); Miller (1953); Umeda & Teranishi (1966); Fujisaki & Kawashima (1968); Fant, Carlson & Granström (1974); Scott (1976); Traunmüller (1981); Di Benedetto (1991; 1994); Hirahara & Kato (1992).

10 The results of the normalization procedures could deviate from the present findings when speech from children and female speakers is included in the data set. Some normalization procedures might perform less strongly, and others could improve because of the differences in vocal tract length.

11 The difference between F3 and F2 was one of the factors used in the intrinsic method of Syrdal & Gopal (1986).

12 The recordings have been made in the AVIP (Bertinetto 2001) and in the API (Crocco, Savy, Cutugno 2002) projects.

13 See Ferrari Disner (1980): vowel normalization procedures are successful only on groups of speakers with a similar phonological vowel system.

14 And Leghorn Italian (Calamai 2001).

15 Macros worked out by Ferrero (1995) and Cioni (2001) were used in the acoustic analysis.

16 Segmentation criteria described in Salza (1991) were adopted.

17 This test statistic compares directly to the F-ratio in ANOVA. All the Hotelling's trace values were displayed in the Appendices.

[18] If Table 3 of § I.1.1 and Table 1 in the Appendices were compared, it could be seen that the parameter set F1-*f*0 x F2 in ERB scale obtained the highest values of Hotelling's F (3328.347) and therefore obtained the highest rank score (42); whereas the opposite happened concerning the parameter set F1-f0 x F3-F2 in Hz, which obtained the lowest value of Hotelling's F and therefore the lowest rank score (1).

[19] R method of classification or resubstitution. The U method was also used, but since the differences between the two methods were always very small, only the values of R method of classification were given in the following pages. The prior probabilities of group membership were assumed not to be equal and were computed from group sizes.

[20] The ranking scores were always computed without duration. Some observations concerning the temporal variable were nevertheless put forward in the following pages (especially in the case of discriminant analyses). Moreover, all the information about the contribution of duration could be found in the Appendices (MANOVAs and discriminant analyses) and in the fifth column of tables 4, 13, 24, 27, 35 (discriminant analyses). As for discriminant analyses run, percentages of correct vowel identification were displayed in the tables, instead of the rank scores, which have been computed separately and due to space limitation were omitted.

[21] As for the highest scores and percentages, they were displayed as far as those reached by F1 x F2 and by F1 x (F2-F1) in Hz.

[22] The intermediate rank scores omitted in Table 3 are the following: 25 (F1-*f*0 x F2-*f*0, Hz and F1-*f*0 x F2-F1, Hz), 24 (F1-*f*0 x F2-F1, Bark), 23 (F1 x F2-F1, Koenig-scale), 22 (F1-*f*0 x F2-F1, Koenig-scale), 21 (F1 x F2-F1, Mel), 20 (F1-*f*0 x F2-F1, Mel), 19 (F1-*f*0 x F2-*f*0, Koenig-scale), 18 (F2-F1 x F3-F2, Hz), 17 (F2-F1 x F3-F2, Bark), 16 (F2-F1 x F3-F2, Koenig-scale), 15 (F1 x F2-F1, ERB), 14 (F1-*f*0 x F2-F1, ERB), 13 (F2-F1 x F3-F2, Mel), 12 (F1 x F2-F1, LogHz), 11 (F1-*f*0 x F2-F1, LogHz), 10 (F1-*f*0 x F3-F2, Bark), 9 (F2-F1 x F3-F2, ERB), 8 (F1 x F2, Koenig-scale), 7 (F1-*f*0 x F2, Koenig-scale), 6 (F1-*f*0 x F3-F2, LogHz), 5 (F2-F1 x F3-F2, LogHz), 4 (F1-f0 x F3-F2, ERB).

[23] The intermediate percentages omitted in Table 4 are the following: 91.4% (F1-f0 x F2-f0 Koenig-scale), 91.6% (F1-f0 x F2-f0 + duration, Koenig-scale); 91.2% (F1-*f*0 x F3-F2, Bark; F1-*f*0 x F3-F2 + duration, Bark); 91.2% (F1 x F2, Koenig-scale; F1 x F2 + duration, Koenig-scale); 91.2% (F1-*f*0 x F2-F1, Koenig-scale), 91.4% (F1-*f*0 x F2-F1 + duration, Koenig-scale); 90.9% (F1-*f*0 x F2-*f*0, Hz), 91.5% (F1-f0 x F2-*f*0 + duration, Hz); 90.9% (F1-*f*0 x F2-F1, Hz), 91.5% (F1-*f*0 x F2-F1 + duration, Hz); 90.9% (F1-*f*0 x F2, Hz), 91.8% (F1-*f*0 x F2 + duration, Hz); 90.7% (F1-*f*0 x F2, Koenig-scale; F1-*f*0 x F2 + duration, Koenig-scale); 88.9% (F1-*f*0 x F3-F2, Mel), 88.6% (F1-*f*0 x F3-F2 + duration, Mel); 88.4% (F1-*f*0 x F3-F2, Koenig-scale), 87.7% (F1-*f*0 x F3-F2 + duration, Koenig-scale); 88.1% (F1-*f*0 x F3-F2, ERB), 88.6% (F1-*f*0 x F3-F2 + duration, ERB); 88% (F1-*f*0 x F3-F2, Hz), 87.7% (F1-*f*0 x F3-F2 + duration, Hz); 87.5% (F1-*f*0 x F3-F2, LogHz), 88.3% (F1-*f*0 x F3-F2 + duration, LogHz); 87.5% (F2-F1 x F3-F2, Bark), 88% (F2-F1 x F3-F2 + duration, Bark); 75.7% (F2-F1 x F3-F2, Hz), 77.1% (F2-F1 x F3-F2 + duration, Hz); 74.6% (F2-F1 x F3-F2, Koenig-scale), 76.9% (F2-F1 x F3-F2 + duration, Koenig-scale).

[24] Apart from: F1-*f*0 x F2-*f*0 (ERB); F1-*f*0 x F3-F2 (Mel); F1-*f*0 x F3-F2 (Koenig-scale); F1-*f*0 x F3-F2 (Hz).

[25] As for /ɛ/, the correct identification percentage reaches 90.5%; as for /o/ it reaches 87.6%; as for /u/ it reaches 87%. The overall correct identification percentage is 92%.

[26] In comparison with the same parameter set in the Hz scale (F1 x F2-F1), high and mid-high front vowels are better discriminated, the opposite is for /a/ and /ɛ/.

[27] The overall correct identification percentage is 76.5%.

[28] The intermediate rank scores omitted in Table 11 are the following: 20 (F1-*f*0 x F2-F1, Bark; F1-*f*0 x F2-F1, Mel; F1-*f*0 x F2-F1, ERB), 19 (F1-*f*0 x F2, Hz), 15 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz; F1-*f*0 x F2-F1, Koenig-scale; F1-*f*0 x F2-*f*0, Koenig-scale), 14 (F1-*f*0 x F3-F2, Bark), 13 (F1 x F2, Koenig-scale), 11 (F2-F1 x F3-F2, Hz; F2-F1 x F3-F2, Bark), 10 (F2-F1 x F3-F2, Koenig-scale), 9 (F1-*f*0 x F2, Koenig-scale), 8 (F2-F1 x F3-F2, Mel), 7 (F1-*f*0 x F3-F2, Mel), 6 (F1-*f*0 x F3-F2, ERB), 4 (F1-*f*0 x F3-F2, LogHz; F1-*f*0 x F3-F2, Koenig-scale).

[29] The intermediate rank scores not included in Table 12 are the following: 40 (F1 x F2, Mel), 39 (F1-*f*0 x F2, Hz), 37 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 36 (F1 x F2-F1, Koenig-scale), 35 (F1 x F2, Bark), 34 (F1 x F2-F1, Bark), 33 (F1 x F2, Koenig-scale), 32 (F1-*f*0 x F2, Mel), 31 (F1-*f*0 x F2-*f*0, Mel), 30 (F1 x F2, ERB), 29 (F1-*f*0 x F2-F1, Koenig-scale), 28 (F1 x F2-F1, Mel), 27 (F1-*f*0 x F2-*f*0, Koenig-scale), 26 (F1 x F2, LogHz), 25 (F1-*f*0 x F2, Bark), 24 (F1-*f*0 x F2-*f*0, Bark), 23 (F1-f0 x F2, Koenig-scale), 22 (F1-*f*0 x F2, ERB), 21 (F1-*f*0 x F2-f0, ERB), 20 (F1-*f*0 x F2-F1, Mel), 19 (F1-*f*0 x F2-F1, Bark), 18 (F1-*f*0 x F2, LogHz), 17 (F1-*f*0 x F2-*f*0, LogHz), 16 (F2-F1 x F3-F2, Bark), 15 (F2-F1 x F3-F2, Hz), 14 (F1 x F2-F1, ERB), 13 (F1-*f*0 x F2-F1, ERB), 12 (F1-*f*0 x F3-F2, Bark), 11 (F1 x F2-F1, LogHz), 10 (F2-F1 x F3-F2, Koenig-scale), 9 (F2-F1 x F3-F2, Mel), 8 (F1-*f*0 x F2-F1, LogHz), 7 (F1-*f*0 x F3-F2, Mel), 6 (F1-*f*0 x F3-F2, Hz), 5 (F1-*f*0 x F3-F2, ERB), 4 (F2-F1 x F3-F2, ERB).

[30] The intermediate percentages not included in Table 13 are the following: 89% (F1-*f*0 x F2, Hz), 89.5% (F1-*f*0 x F2 + duration, Hz); 89% (F1 x F2, Mel), 89.7% (F1 x F2 + duration, Mel); 89% (F1-*f*0 x F2-F1, Mel), 89.5% (F1-*f*0 x F2-F1 + duration, Mel); 88.9% (F1-*f*0 x F2-F1, Hz), 89.3% (F1-*f*0 x F2-F1 + duration, Hz); 88.9% (F1-*f*0 x F2-F1, Koenig-scale), 89% (F1-*f*0 x F2-F1 + duration, Koenig-scale); 88.8% (F1-*f*0 x F2, Bark), 89.2% (F1-*f*0 x F2 + duration, Bark); 88.7% (F1-*f*0 x F2, Koenig-scale), 88.8% (F1-*f*0 x F2 + duration, Koenig-scale); 88.6% (F1-*f*0 x F2, LogHz), 89% (F1-*f*0 x F2 + duration, LogHz); 88.6% (F1-*f*0 x F2, ERB), 89.5% (F1-*f*0 x F2 + duration, ERB); 88.6% (F1-*f*0 x F2-*f*0, Koenig-scale), 89.2% (F1-*f*0 x F2-*f*0 + duration, Koenig-scale); 88.4% (F1-*f*0 x F2-F1, ERB), 89.4% (F1-*f*0 x F2-F1 + duration, ERB); 88.2% (F1-*f*0 x F2-*f*0, ERB), 89.1% (F1-*f*0 x F2-*f*0 + duration, ERB); 87.9% (F1-*f*0 x F2-*f*0, LogHz), 88.6% (F1-*f*0 x F2-*f*0 + duration, LogHz); 88.9% (F1-*f*0 x F2-*f*0, Hz), 89.3% (F1-*f*0 x F2-f0 + duration, Hz); 87.9% (F1-*f*0 x F2-*f*0, Bark), 88.6% (F1-*f*0 x F2-*f*0 + duration, Bark); 87.5% (F1-*f*0 x F2-F1, LogHz), 88.5% (F1-*f*0 x F2-F1 + duration, LogHz); 87.3% (F1-*f*0 x F3-F2, Bark), 88% (F1-*f*0 x F3-F2 + duration, Bark); 87.3% (F1-*f*0 x F3-F2, Mel), 87.2% (F1-*f*0 x F3-F2 + duration, Mel); 87.2% (F1-*f*0 x F2-F1, Bark), 88% (F1-*f*0 x F2-F1 + duration, Bark); 86.8% (F1-*f*0 x F3-F2, ERB), 87% (F1-*f*0 x F3-F2 + duration, ERB); 86.6% (F1-*f*0 x F3-F2, LogHz), 86.8% (F1-*f*0 x F3-F2 + duration, LogHz); 86.5% (F1-*f*0 x F3-F2, Hz), 87% (F1-*f*0 x F3-F2 + duration, Hz); 86.1% (F1-*f*0 x F3-F2, Koenig-scale), 85.7% (F1-*f*0 x F3-F2 + duration, Koenig-scale); 83.7% (F2-F1 x F3-F2, Bark), 84.6% (F2-F1 x F3-F2 + duration, Bark); 70.9% (F2-F1 x F3-F2, Mel; F2-F1 x F3-F2 + duration, Mel); 70.3% (F2-F1 x F3-F2, Hz), 70.4% (F2-F1 x F3-F2 + duration, Hz).

[31] Apart from: F1 x F2 (Koenig-scale); F1-*f*0 x F3-F2 (Mel); F1-*f*0 x F3-F2 (Koenig-scale).

[32] The intermediate rank scores not included in Table 18 are the following: 35 (F1 x F2, Koenig-scale; F1-*f*0 x F2, Mel; F1 x F2, ERB), 33 (F1 x F2, Mel; F1 x F2-F1, Bark), 32 (F1-*f*0 x F2, Hz), 31 (F1 x F2, LogHz), 29 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 28 (F1-*f*0 x F2-*f*0, Mel), 27 (F1 x F2-F1, ERB), 26 (F1-*f*0 x F2-F1, Koenig-scale), 25 (F1 x F2-F1, LogHz), 24 (F1-*f*0 x F2, Bark), 23 (F1-*f*0 x F2-F1, Mel), 22 (F1-*f*0 x F2-*f*0, Koenig-scale), 21 (F1-*f*0 x F2, Koenig-scale), 20 (F1-*f*0 x

F2, ERB), 19 (F1-*f*0 x F2-f0, Bark), 18 (F1-*f*0 x F2-*f*0, ERB), 17 (F1-*f*0 x F2, LogHz), 16 (F1-*f*0 x F2-*f*0, LogHz), 15 (F1-*f*0 x F2-F1, ERB), 14 (F1-*f*0 x F2-F1, Bark), 13 (F1-f0 x F3-F2, Bark), 11 (F2-F1 x F3-F2, Bark; F1-f0 x F2-F1, logHz), 9 (F2-F1 x F3-F2, Hz; F1-*f*0 x F3-F2, Mel), 8 (F1-*f*0 x F3-F2, ERB), 6 (F2-F1 x F3-F2, Mel; F1-*f*0 x F3-F2, Hz), 5 (F2-F1 x F3-F2, Koenig-scale), 4 (F1-*f*0 x F3-F2, LogHz).

[33] The intermediate rank scores not included in Table 23 are the following: 36 (F1 x F2-F1, Koenig-scale), 34 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 33 (F1-*f*0 x F2-*f*0, Mel), 32 (F1 x F2, Koenig-scale), 31 (F1-*f*0 x F2-F1, Koenig-scale), 30 (F1-*f*0 x F2-*f*0, Bark), 29 (F1-*f*0 x F2-*f*0, ERB), 28 (F1 x F2-F1, Mel), 27 (F1-*f*0 x F2-*f*0, Koenig-scale), 26 (F1-*f*0 x F2, Hz), 25 (F1-*f*0 x F2, Mel), 24 (F1-*f*0 x F2, Bark), 23 (F1-*f*0 x F2, ERB), 22 (F1 x F2-F1, Bark), 21 (F1-*f*0 x F2-*f*0, LogHz), 20 (F1-*f*0 x F2, Koenig-scale), 19 (F1-*f*0 x F2-F1, Mel), 18 (F1-*f*0 x F2, LogHz), 17 (F1 x F2-F1, ERB), 16 (F1-*f*0 x F2-F1, Bark), 15 (F1 x F2-F1, LogHz), 14 (F2-F1 x F3-F2, Hz), 13 (F1-*f*0 x F2-F1, ERB), 12 (F2-F1 x F3-F2, Koenig-scale), 11 (F1-*f*0 x F2-F1, LogHz), 10 (F2-F1 x F3-F2, Mel), 9 (F2-F1 x F3-F2, Bark), 8 (F1-*f*0 x F3-F2, ERB), 7 (F1-*f*0 x F3-F2, Bark), 6 (F1-*f*0 x F3-F2, Mel), 5 (F1-*f*0 x F3-F2, LogHz), 4 (F2-F1 x F3-F2, ERB).

[34] The intermediate percentages not included in Table 24 are the following: 96.8% (F1 x F2, Koenig-scale), 97.8% (F1 x F2 + duration, Koenig-scale); 95.5% (F1-*f*0 x F2-*f*0, Hz), 96.2% (F1-*f*0 x F2-*f*0 + duration, Hz); 95.5% (F1-*f*0 x F2-*f*0, LogHz), 96.5% (F1-*f*0 x F2-*f*0 + duration, LogHz); 95.5% (F1-*f*0 x F2-F1, Hz), 96.2% (F1-*f*0 x F2-F1 + duration, Hz); 95.5% (F1-*f*0 x F2, Bark), 96.2% (F1-*f*0 x F2 + duration, Bark); 95.5% (F1-*f*0 x F2-*f*0, Mel), 95.9% (F1-*f*0 x F2-*f*0 + duration, Mel); 95.5% (F1-*f*0 x F2, Koenig-scale), 96.2% (F1-*f*0 x F2 + duration, Koenig-scale); 95.2% (F1-*f*0 x F2-F1, ERB), 96.2% (F1-*f*0 x F2-F1 + duration, ERB); 95.2% (F1-*f*0 x F2-*f*0, ERB), 95.9% (F1-*f*0 x F2-*f*0 + duration, ERB); 95.2% (F1-*f*0 x F2, LogHz), 96.2% (F1-*f*0 x F2 + duration, LogHz); 95.2% (F1-*f*0 x F2-*f*0, Bark), 95.5% (F1-*f*0 x F2-*f*0 + duration, Bark); 95.2% (F1-*f*0 x F2-F1, Mel), 96.2% (F1-*f*0 x F2-F1 + duration, Mel); 95.2% (F1-*f*0 x F2, Mel), 96.2% (F1-*f*0 x F2 + duration, Mel); 95.2% (F1-*f*0 x F2, ERB), 96.5% (F1-*f*0 x F2 + duration, ERB); 95.2% (F1-*f*0 x F2-*f*0, Koenig-scale), 95.5% (F1-*f*0 x F2-*f*0 + duration, Koenig-scale); 95.2% (F1-*f*0 x F2-F1, Bark), 96.2% (F1-*f*0 x F2-F1 + duration, Bark); 94.9% (F1-*f*0 x F2-F1, LogHz), 96.2% (F1-*f*0 x F2-F1 + duration, LogHz); 94.9% (F1-*f*0 x F2-F1, Koenig-scale), 96.2% (F1-*f*0 x F2-F1 + duration, Koenig-scale); 94.9% (F1-*f*0 x F2, Hz), 96.2% (F1-*f*0 x F2 + duration, Hz); 90.4% (F1-*f*0 x F3-F2, Hz), 91.7% (F1-*f*0 x F3-F2 + duration, Hz); 90.1% (F1-*f*0 x F3-F2, Mel), 91.4% (F1-*f*0 x F3-F2 + duration, Mel); 90.1% (F1-*f*0 x F3-F2, Bark), 91.4% (F1-*f*0 x F3-F2 + duration, Bark); 89.8% (F1-*f*0 x F3-F2, Koenig-scale), 91.4% (F1-*f*0 x F3-F2 + duration, Koenig-scale); 89.8% (F1-*f*0 x F3-F2, ERB), 91.1% (F1-*f*0 x F3-F2 + duration, ERB); 88.2% (F1-*f*0 x F3-F2, LogHz), 89.2% (F1-*f*0 x F3-F2 + duration, LogHz); 74.5% (F2-F1 x F3-F2, ERB), 74.8% (F2-F1 x F3-F2 + duration, ERB); 74.2% (F2-F1 x F3-F2, Bark), 75.8% (F2-F1 x F3-F2 + duration, Bark); 73.9% (F2-F1 x F3-F2, Hz), 75.8% (F2-F1 x F3-F2 + duration, Hz).

[35] Apart from: F1 x F2-F1 (LogHz); F1 x F2-F1 (ERB); F2-F1 x F3-F2 (LogHz).

[36] The intermediate rank scores not included in Table 25 are the following: 35 (F1 x F2, Koenig-scale), 34 (F1 x F2-F1, Mel), 32 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 30 (F1-*f*0 x F2-*f*0, Mel; F1 x F2-F1, ERB), 29 (F1 x F2-F1, LogHz), 28 (F1 x F2-F1, Bark), 27 (F1-*f*0 x F2, Bark), 25 (F1-*f*0 x F2-f0, LogHz; F1-*f*0 x F2-*f*0, Bark), 23 (F1-*f*0 x F2-*f*0, ERB; F1-*f*0 x F2, Koenig-scale), 22 (F1-*f*0 x F2-F1, Koenig-scale), 21 (F1-*f*0 x F2-*f*0, Koenig-scale), 20 (F1-*f*0 x F2, Mel), 18 (F1-*f*0 x F2, Hz; F1-*f*0 x F2, ERB), 17 (F1-*f*0 x F2-F1, Mel), 16 (F1-*f*0 x F2, LogHz), 15 (F1-*f*0 x F2-F1, Bark), 14 (F1-*f*0 x F2-F1, ERB), 13 (F1-*f*0 x F2-F1, LogHz), 12 (F1-*f*0 x

F3-F2, Bark), 10 (F1-*f*0 x F3-F2, ERB; F1-*f*0 x F3-F2, Mel), 9 (F2-F1 x F3-F2, Hz), 7 (F2-F1 x F3-F2, Bark; F1-*f*0 x F3-F2, Hz), 6 (F2-F1 x F3-F2, Koenig-scale), 5 (F1-*f*0 x F3-F2, logHz), 3 (F2-F1 x F3-F2, Mel).

[37] The intermediate rank scores not included in Table 26 are the following: 36 (F1 x F2, Koenig-scale), 35 (F1 x F2-F1, Koenig-scale), 34 (F1-*f*0 x F2-*f*0, ERB), 33 (F1-*f*0 x F2-*f*0, Mel), 32 (F1-*f*0 x F2-*f*0, Bark), 31 (F1 x F2-F1, Mel), 30 (F1-*f*0 x F2, ERB), 29 (F1-*f*0 x F2, Bark), 28 (F1-*f*0 x F2, Mel), 27 (F1-*f*0 x F2-*f*0, Koenig-scale), 26 (F1 x F2-F1, Bark), 25 (F1-*f*0 x F2-*f*0, LogHz), 23 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 22 (F1-*f*0 x F2-F1, Koenig-scale), 21 (F2-F1 x F3-F2, Hz), 20 (F1-*f*0 x F2, Koenig-scale), 19 (F1-*f*0 x F2, LogHz), 18 (F1-*f*0 x F2, Hz), 17 (F1-*f*0 x F2-F1, Mel), 16 (F2-F1 x F3-F2, Koenig-scale), 15 (F1 x F2-F1, ERB), 14 (F1-*f*0 x F2-F1, Bark), 13 (F1 x F2-F1, LogHz), 12 (F1-*f*0 x F2-F1, ERB), 11 (F2-F1 x F3-F2, Mel), 10 (F2-F1 x F3-F2, Bark), 9 (F1-*f*0 x F2-F1,LogHz), 8 (F2-F1 x F3-F2, ERB), 7 (F2-F1 x F3-F2, LogHz), 6 (F1-*f*0 x F3-F2, LogHz), 5 (F1-*f*0 x F3-F2, ERB), 4 (F1-*f*0 x F3-F2, Bark).

[38] The intermediate percentages not included in Table 27 are the following: 77% (F1-*f*0 x F2, LogHz), 78.9% (F1-*f*0 x F2 + duration, LogHz); 77% (F1-*f*0 x F2-*f*0, Bark), 78.5% (F1-*f*0 x F2-*f*0 + duration, Bark); 76.3% (F1-*f*0 x F2-F1, Mel), 77.3% (F1-*f*0 x F2-F1 + duration, Mel); 76.3% (F1-*f*0 x F2, Mel), 77.3% (F1-*f*0 x F2 + duration, Mel); 76% (F1-*f*0 x F2, ERB), 77.3% (F1-*f*0 x F2 + duration, ERB); 76% (F1-*f*0 x F2, Bark), 77.3% (F1-*f*0 x F2 + duration, Bark); 75.4% (F1-*f*0 x F2-*f*0, Koenig-scale), 76.3% (F1-*f*0 x F2-*f*0 + duration, Koenig-scale); 75.1% (F1-*f*0 x F2-F1, Koenig-scale), 76.7% (F1-*f*0 x F2-F1 + duration, Koenig-scale); 75.1% (F1-*f*0 x F2, Koenig-scale), 74.8% (F1-*f*0 x F2 + duration, Koenig-scale); 74.8% (F1-*f*0 x F2, Hz), 76% (F1-*f*0 x F2 + duration, Hz); 74.1% (F1-*f*0 x F2-F1, Bark), 77.6% (F1-*f*0 x F2-F1 + duration, Bark); 71.3% (F1-*f*0 x F3-F2, ERB), 73.7% (F1-*f*0 x F3-F2 + duration, ERB); 71.3% (F1-*f*0 x F3-F2, LogHz; F1-*f*0 x F3-F2 + duration, LogHz), 70.2% (F1-*f*0 x F3-F2, Hz), 74.4% (F1-*f*0 x F3-F2 + duration, Hz); 70.2% (F1-*f*0 x F3-F2, Mel), 74.7% (F1-*f*0 x F3-F2 + duration, Mel); 70.2% (F1-*f*0 x F3-F2, Bark), 74% (F1-*f*0 x F3-F2 + duration, Bark); 69.9% (F1-*f*0 x F3-F2, Koenig-scale), 72.3% (F1-*f*0 x F3-F2 + duration, Koenig-scale); 68.2% (F2-F1 x F3-F2, Hz), 69.4% (F2-F1 x F3-F2 + duration, Hz); 68.2% (F2-F1 x F3-F2, Mel), 69.4% (F2-F1 x F3-F2 + duration, Mel); 66.4% (F2-F1 x F3-F2, Koenig-scale), 66.7% (F2-F1 x F3-F2 + duration, Koenig-scale).

[39] Many studies focussing on the differences between read and semi-spontaneous speech are now available: see for instance Ladefoged, Kameny & Brackenridge (1976); ESCA (1991); Llisterri & Poch-Olivé (1992); Simpson & Pätzold (1996); Barry & Andreeva (2001).

[40] In three cases duration do not improve vowels classifications: F1 x (F2-F1) in LogHz, F1x (F2-F1) in ERB, (F1-*f*0) x F2 in Koenig-scale.

[41] The intermediate rank scores not included in Table 28 are the following: 34 (F1 x F2-F1, Koenig-scale), 33 (F1-*f*0 x F2-*f*0, ERB), 32 (F1-*f*0 x F2-*f*0, Mel), 31 (F1-*f*0 x F2-*f*0, Bark), 29 (F1 x F2-F1, Bark; F1 x F2-F1, LogHz), 28 (F1-*f*0 x F2-*f*0, LogHz), 27 (F1 x F2-F1, ERB), 25 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 22 (F1-*f*0 x F2-F1, ERB; F1-*f*0 x F2, Mel; F1-*f*0 x F2, ERB), 21 (F1-*f*0 x F2, Bark), 20 (F1-*f*0 x F2-*f*0, Koenig-scale), 19 (F1-*f*0 x F2, LogHz), 17 (F1-*f*0 x F2-F1, Koenig-scale; F1-*f*0 x F2-F1, Mel), 15 (F1-*f*0 x F2, Koenig-scale; F1-*f*0 x F2-F1, LogHz), 14 (F1-*f*0 x F2, Hz), 13 (F1-*f*0 x F2-F1, Bark), 12 (F2-F1 x F3-F2, Hz), 11 (F2-F1 x F3-F2, Koenig-scale), 10 (F1-*f*0 x F3-F2, LogHz), 8 (F2-F1 x F3-F2, Mel; F1-*f*0 x F3-F2, ERB), 7 (F2-F1 x F3-F2, Bark), 6 (F1-*f*0 x F3-F2, Bark), 5 (F1-*f*0 x F3-F2, Mel), 3 (F1-*f*0 x F3-F2, Hz).

[42] F1 x F2 in Bark, F1 x (F2-F1) in LogHz, (F1-*f*0) x (F2-F1) in Mel, (F2-F1) x (F3-F2) in Bark, (F2-F1) x (F3-F2) in LogHz.

[43] The intermediate rank scores not included in Table 33 are the following: 33 (F1-*f*0 x F2-*f*0, Mel), 32 (F1 x F2-F1, LogHz), 28 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz; F1 x F2-F1, ERB; F1 x F2-F1, Bark), 26 (F1-*f*0 x F2-*f*0, Bark; F1-*f*0 x F2-*f*0, ERB), 25 (F1-*f*0 x F2-*f*0, LogHz), 24 (F1-*f*0 x F2, Bark), 23 (F1-*f*0 x F2, Mel), 22 (F1-*f*0 x F2-*f*0, Koenig-scale), 21 (F1-*f*0 x F2, ERB), 20 (F1-*f*0 x F2-F1, Koenig-scale), 19 (F1-*f*0 x F2, Koenig-scale), 18 (F1-*f*0 x F2-F1, ERB), 17 (F1-*f*0 x F2, LogHz), 16 (F1-*f*0 x F2-F1, Mel), 15 (F1-*f*0 x F2, Hz), 13 (F1-*f*0 x F2-F1, Bark; F1-*f*0 x F2-F1, LogHz), 12 (F2-F1 x F3-F2, Hz), 10 (F1-*f*0 x F3-F2, Bark; F1-*f*0 x F3-F2, ERB), 9 (F2-F1 x F3-F2, Koenig-scale), 7 (F1-*f*0 x F3-F2, Mel; F1-*f*0 x F3-F2, LogHz), 6 (F2-F1 x F3-F2, Bark), 5 (F2-F1 x F3-F2, Mel), 4 (F1-*f*0 x F3-F2, Hz).

[44] The intermediate rank scores not included in Table 34 are the following: 22 (F1-*f*0 x F2-F1, Koenig-scale), 21 (F1-*f*0 x F2-F1, Mel), 20 (F1 x F2-F1, Koenig-scale), 19 (F1-*f*0 x F2-F1, Bark), 18 (F1 x F2-F1, Mel), 17 (F1-*f*0 x F2-F1, ERB), 16 (F1 x F2-F1, Bark), 15 (F1-*f*0 x F2-F1, LogHz), 14 (F1 x F2-F1, ERB), 13 (F2-F1 x F3-F2, Hz), 12 (F1 x F2-F1, LogHz), 11 (F1-*f*0 x F3-F2, Mel), 10 (F1-*f*0 x F3-F2, ERB), 9 (F1-*f*0 x F3-F2, Bark), 8 (F2-F1 x F3-F2, Koenig-scale), 7 (F1-*f*0 x F3-F2, Koenig-scale), 6 (F1-*f*0 x F3-F2, Hz), 5 (F1-*f*0 x F3-F2, LogHz), 4 (F2-F1 x F3-F2, Mel).

[45] The intermediate percentages not included in Table 35 are the following: 90.4% (F1 x F2, Koenig-scale), 90.1% (F1 x F2 + duration, Koenig-scale); 90.1% (F1-*f*0 x F2, Bark), 89.8% (F1-*f*0 x F2 + duration, Bark); 90.1% (F1 x F2, LogHz), 90.4% (F1 x F2 + duration, LogHz); 89.8% (F1 x F2, Mel), 91.6% (F1 x F2 + duration, Mel); 88.9% (F1-*f*0 x F2-*f*0, Bark), 88.9% (F1-*f*0 x F2-*f*0 + duration, Bark); 88.5% (F1 x F2, Bark), 87.6% (F1 x F2 + duration, Bark); 88.2% (F1 x F2-F1, Mel), 89.2% (F1 x F2-F1 + duration, Mel); 87.6% (F1-f0 x F2-F1, ERB), 88.9% (F1-f0 x F2-F1 + duration, ERB); 87.6% (F1-*f*0 x F2-F1, LogHz), 87.9% (F1-*f*0 x F2-F1 + duration, LogHz); 87.6% (F1 x F2-F1, Koenig-scale; F1 x F2-F1 + duration, Koenig-scale); 87% (F1-*f*0 x F3-F2, Mel), 87.3% (F1-*f*0 x F3-F2, Mel); 86.6% (F1-f0 x F3-F2, ERB; F1-*f*0 x F3-F2 + duration, ERB); 86.6% (F1-*f*0 x F3-F2, LogHz), 86.6% (F1-*f*0 x F3-F2 + duration, logHz); 85.8% (F1-*f*0 x F2-F1, Bark), 87.3% (F1-*f*0 x F2-F1 + duration, Bark); 85.4% (F1-*f*0 x F3-F2, Koenig-scale), 86.6% (F1-*f*0 x F3-F2 + duration, Koenig-scale); 85.1% (F1 x F2-F1, ERB), 86.7% (F1 x F2-F1 + duration, ERB); 84.8% (F1-*f*0 x F3-F2, Hz), 87% (F1-*f*0 x F3-F2 + duration, Hz); 84.8% (F1 x F2-F1, LogHz), 86.1% (F1 x F2-F1 + duration, LogHz); 84.2% (F1 x F2-F1, Bark), 86.1% (F1 x F2-F1 + duration, Bark); 83.2% (F1-*f*0 x F3-F2, Bark), 86.6% (F1-*f*0 x F3-F2 + duration, Bark); 76.7% (F2-F1 x F3-F2, Hz), 80.1% (F2-F1 x F3-F2 + duration, Hz); 74.5% (F2-F1 x F3-F2, Mel), 76.7% (F2-F1 x F3-F2 + duration, Mel); 73.9% (F2-F1 x F3-F2, Koenig-scale), 75.8% (F2-F1 x F3-F2 + duration, Koenig-scale).

[46] The overall recognition rate gets 87.3%.

[47] The intermediate rank scores not included in Table 37 are the following: 20 (F1 x F2-F1, Mel), 19 (F1 x F2-F1, Koenig-scale), 18 (F1-*f*0 x F2-F1, ERB), 16 (F1-f0 x F2-F1, Bark; F1-*f*0 x F2-F1, LogHz), 15 (F1-*f*0 x F3-F2, Mel), 14 (F1 x F2-F1, ERB), 12 (F1 x F2-F1, Bark; F1-*f*0 x F3-F2, ERB), 11 (F1 x F2-F1, LogHz), 8 (F2-F1 x F3-F2, Hz; F1-*f*0 x F3-F2, Koenig-scale; F1-*f*0 x F3-F2, LogHz), 7 (F1-*f*0 x F3-F2, Bark), 6 (F1-*f*0 x F3-F2, Hz), 5 (F2-F1 x F3-F2, Koenig-scale), 4 (F2-F1 x F3-F2, Mel).

[48] In the Florence sample the confusion involves the contiguous vowel /e/, not the contiguous vowel /a/.

[49] The intermediate rank scores not included in Table 42 are the following: 26 (F1-*f*0 x F2, Bark), 24 (F1-*f*0 x F2-f0, Hz; F1-*f*0 x F2-F1, Hz), 23 (F1 x F2-F1, Koenig-scale), 22 (F1-*f*0 x F2, Koenig-scale), 21 (F1-*f*0 x F2, Hz), 20 (F1-*f*0 x F2-F1, Koenig-scale), 18 (F1 x F2-F1, ERB; F1 x F2-F1, Bark), 15 (F1 x F2-F1,

LogHz; F1-*f*0 x F2-F1, ERB; F1-*f*0 x F2-F1, Mel); 14 (F1-*f*0 x F2-F1, LogHz), 13 (F1-*f*0 x F2-F1, Bark), 10 (F2-F1 x F3-F2, Hz; F1-f0 x F3-F2, ERB; F1-*f*0 x F3-F2, Mel), 9 (F1-*f*0 x F3-F2, LogHz), 8 (F2-F1 x F3-F2, Koenig-scale), 7 (F1-*f*0 x F3-F2, Bark), 6 (F2-F1 x F3-F2, Mel), 3 (F2-F1 x F3-F2, Bark; F1-*f*0 x F3-F2, Koenig-scale).

[50] As for the Pisa sample, formant data are displayed in Figure 5.

[51] The normalization procedures we applied should not produce procedural artifacts like the ones described in Ferrari Disner (1980) when comparing the normalized vowels of one language with the (independently) normalized vowels of another language.

[52] In the unstressed sample the absolute highest score was reached by F1 x F2 in Mel transform, but we believe that this finding does not question the validity of our observation.

[53] In the stressed condition, (F1-*f*0) x F2 in ERB transform scored 42, whereas in the unstressed condition scored 22. In the unstressed condition, F1 x F2 in Hz scored 41, whereas in the stressed condition scored 27.

[54] See Heid (1997:761-762); Adank, van Heuven, van Hout (1999:1596).

[55] See Hillenbrand & Gayvert (1993:698); Ferrero (1994:11); Cosi, Ferrero, Vagges (1995:139); Syrdal (1985:130).

[56] See Tables 1, 2, 5, 6, 7 in the Appendices.

[57] "Adding vowel duration measures resulted in consistent but fairly modest improvements in classification accuracy, and including two samples of formant pattern produced large improvements in category separability" (Hillenbrand *et al*. 1995:3109). See also Weenink (2001:120): "Including dynamics improves the classification process. The classification results for the dynamic spectra are always better than those for the corresponding static spectra".

[58] To be more precise: the arithmetic mean of three different measurements taken in the steady state, as pointed out in § 0.2.

[59] Many studies showed that spectral dynamics of vowel realizations are useful in vowel classification: see for instance Kuwabara (1985); Di Benedetto (1989); Huang (1992).

[60] The intermediate Hotelling's F values omitted in Table 1 are the following: 3059.652 (F1-*f*0 x F2-*f*0, Hz), 2078.326 (F1-*f*0 x F2-*f*0 + duration, Hz); 3059.652 (F1-*f*0 x F2-F1, Hz), 2078.326 (F1-*f*0 x F2-F1 + duration, Hz); 2931.793 (F1-*f*0 x F2-F1, Bark), 2000.053 (F1-*f*0 x F2-F1 + duration, Bark); 2819.295 (F1 x F2-F1, Koenig-scale), 1903.053 (F1 x F2-F1 + duration, Koenig-scale); 2818.082 (F1-*f*0 x F2-F1, Koenig-scale), 1904.294 (F1-*f*0 x F2-F1 + duration, Koenig-scale); 2794.103 (F1 x F2-F1, Mel), 1898.551 (F1 x F2-F1 + duration, Mel); 2771.729 (F1-*f*0 x F2-F1, Mel), 1886.259 (F1-*f*0 x F2-F1 + duration, Mel); 2698.890 (F1-*f*0 x F2-*f*0, Koenig-scale), 1830.944 (F1-*f*0 x F2-f0 + duration, Koenig-scale); 2531.620 (F2-F1 x F3-F2, Hz), 1726.790 (F2-F1 x F3-F2 + duration, Hz); 2496.011 (F2-F1 x F3-F2, Bark), 1699.567 (F2-F1 x F3-F2 + duration, Bark); 2365.393 (F2-F1 x F3-F2, Koenig-scale), 1606.439 (F2-F1 x F3-F2 + duration, Koenig-scale); 2330.935 (F1 x F2-F1, ERB), 1578.470 (F1 x F2-F1 + duration, ERB); 2324.456 (F1-*f*0 x F2-F1, ERB), 1575.454 (F1-*f*0 x F2-F1 + duration, ERB); 2133.421 (F2-F1 x F3-F2, Mel), 1460.663 (F2-F1 x F3-F2 + duration, Mel); 2068.659 (F1 x F2-F1, LogHz), 1397.252 (F1 x F2-F1 + duration, LogHz); 2059.359 (F1-*f*0 x F2-F1, LogHz), 1392.009 (F1-*f*0 x F2-F1 + duration, LogHz); 1764.208 (F1-*f*0 x F3-F2, Bark), 1182.449 (F1-*f*0 x F3-F2 + duration, Bark); 1662.537 (F2-F1 x F3-F2, ERB), 1137.293 (F2-F1 x F3-F2 + duration, ERB); 1585.523 (F1 x F2, Koenig-scale), 1067.492 (F1 x F2 + duration, Koenig-scale); 1578.102 (F1-*f*0 x F2, Koenig-scale), 1060.949 (F1-*f*0 x F2 + duration, Koenig-scale); 1394.888 (F1-*f*0 x F3-F2, LogHz), 931.945 (F1-*f*0 x F3-F2 + duration, LogHz); 1373.709 (F2-F1 x F3-F2, LogHz),

938.684 (F2-F1 x F3-F2 + duration, LogHz); 1329.288 (F1-*f*0 x F3-F2, ERB), 888.650 (F1-*f*0 x F3-F2 + duration, ERB).

[61] The intermediate Hotelling's F values omitted in Table 2 are the following: 1857.844 (F1 x F2, Mel), 1260.952 (F1 x F2 + duration, Mel); 1847.402 (F1-*f*0 x F2, Hz), 1256.290 (F1-*f*0 x F2 + duration, Hz); 1844.858 (F1-*f*0 x F2-*f*0, Hz), 1252.267 (F1-*f*0 x F2-*f*0 + duration, Hz); 1844.858 (F1-*f*0 x F2-F1, Hz), 1252.267 (F1-*f*0 x F2-F1 + duration, Hz); 1841.412 (F1 x F2-F1, Koenig-scale), 1250.074 (F1 x F2-F1 + duration, Koenig-scale); 1805.004 (F1 x F2, Bark), 1226.327 (F1 x F2 + duration, Bark); 1802.518 (F1 x F2-F1, Bark), 1224.952 (F1 x F2-F1 + duration, Bark); 1775.627 (F1 x F2, Koenig-scale), 1204.486 (F1 x F2 + duration, Koenig-scale); 1765.695 (F1-*f*0 x F2, Mel), 1204.404 (F1-*f*0 x F2 + duration, Mel); 1753.418 (F1-*f*0 x F2-*f*0, Mel), 1193.740 (F1-*f*0 x F2-*f*0 + duration, Mel); 1741.444 (F1 x F2, ERB), 1182.980 (F1 x F2 + duration, ERB); 1710.109 (F1-*f*0 x F2-F1, Koenig-scale), 1164.760 (F1-*f*0 x F2-F1 + duration, Koenig-scale); 1694.410 (F1 x F2-F1, Mel), 1150.340 (F1 x F2-F1 + duration, Mel); 1683.042 (F1-*f*0 x F2-*f*0, Koenig-scale), 1145.261 (F1-*f*0 x F2-*f*0 + duration, Koenig-scale); 1669.902 (F1 x F2, LogHz), 1133.636 (F1 x F2 + duration, LogHz); 1658.747 (F1-*f*0 x F2, Bark), 1133.078 (F1-*f*0 x F2 + duration, Bark); 1652.050 (F1-*f*0 x F2-*f*0, Bark), 1127.469 (F1-*f*0 x F2-*f*0 + duration, Bark); 1651.334 (F1-*f*0 x F2, Koenig-scale), 1126.157 (F1-*f*0 x F2 + duration, Koenig-scale); 1650.006 (F1-*f*0 x F2, ERB), 1124.866 (F1-*f*0 x F2 + duration, ERB); 1621.645 (F1-*f*0 x F2-*f*0, ERB), 1103.382 (F1-*f*0 x F2-*f*0 + duration, ERB); 1603.842 (F1-*f*0 x F2-F1, Mel), 1093.094 (F1-*f*0 x F2-F1 + duration, Mel); 1585.884 (F1-*f*0 x F2-F1, Bark) 1081.185 (F1-*f*0 x F2-F1 + duration, Bark); 1547.469 (F1-*f*0 x F2, LogHz), 1049.804 (F1-*f*0 x F2 + duration, LogHz); 1509.631 (F1-*f*0 x F2-*f*0, LogHz), 1021.738 (F1-*f*0 x F2-*f*0 + duration, LogHz); 1458.440 (F2-F1 x F3-F2, Bark), 996.900 (F2-F1 x F3-F2 + duration, Bark); 1409.762 (F2-F1 x F3-F2, Hz), 959.295 (F2-F1 x F3-F2 + duration, Hz); 1405.251 (F1 x F2-F1, ERB), 955.733 (F1 x F2-F1 + duration, ERB); 1318.881 (F1-*f*0 x F2-F1, ERB), 900.750 (F1-*f*0 x F2-F1 + duration, ERB); 1291.699 (F1-*f*0 x F3-F2, Bark), 882.715 (F1-*f*0 x F3-F2 + duration, Bark); 1260.146 (F1 x F2-F1, LogHz), 857.191 (F1 x F2-F1 + duration, LogHz); 1260.001 (F2-F1 x F3-F2, Koenig-scale), 862.804 (F2-F1 x F3-F2 + duration, Koenig-scale); 1187.391 (F2-F1 x F3-F2, Mel), 809.974 (F2-F1 x F3-F2 + duration, Mel); 1146.905 (F1-*f*0 x F2-F1, LogHz), 781.397 (F1-*f*0 x F2-F1 + duration, LogHz); 1092.291 (F1-*f*0 x F3-F2, Mel), 747.678 (F1-*f*0 x F3-F2, Mel); 1060.238 (F1-*f*0 x F3-F2, Hz), 724.903 (F1-*f*0 x F3-F2 + duration, Hz); 1057.369 (F1-*f*0 x F3-F2, ERB), 722.325 (F1-*f*0 x F3-F2 + duration, ERB); 982.537 (F2-F1 x F3-F2, ERB), 670.221 (F2-F1 x F3-F2 + duration, ERB).

[62] The intermediate Hotelling's F values omitted in Table 3 are the following: 89.2% (F1-*f*0 x F2-*f*0, Hz); 89.7% (F1 x F2, Koenig-scale); 89.4% (F1-*f*0 x F3-F2, Koenig-scale); 89.2% (F1-*f*0 x F2-F1, Hz); 88.8% (F1-*f*0 x F2, Hz); 88.8% (F1-*f*0 x F2-*f*0, Koenig-scale); 88.4% (F1-*f*0 x F2, Koenig-scale); 72.8% (F3-F2 x F2-F1, Mel); 72.8% (F3-F2 x F2-F1, Hz; F3-F2 x F2-F1, Bark).

[63] The intermediate Hotelling's F values omitted in Table 4 are the following: 90.6% (F1-*f*0 x F2, Bark); 90.5% (F1-*f*0 x F2-F1, Hz); 90.5% (F1-*f*0 x F2, Hz; F1-*f*0 x F2, Koenig-scale; F1-*f*0 x F2-*f*0, Hz); 90.3% (F1-*f*0 x F2-*f*0, Mel; F1-*f*0 x F2-*f*0, Koenig-scale); 90.3% (F1-*f*0 x F2-*f*0, Bark); 89.9% (F1 x F2, LogHz); 89.9% (F1 x F2, ERB; F1-*f*0 x F2, ERB); 89.8% (F1-*f*0 x F2-F1, Bark; F1 x F2-F1, ERB); 89.6% (F1-*f*0 x F2-F1, ERB); 89.5% (F1-*f*0 x F3-F2, Bark); 89.2% (F1 x F2-F1, LogHz); 88.9% (F1-*f*0 x F2, LogHz); 88.6% (F1-*f*0 x F2-*f*0, LogHz); 88.4% (F1-*f*0 x F2-F1, LogHz); 87.8% (F1-*f*0 x F3-F2, Hz); 87.7% (F1-*f*0 x F2-*f*0, ERB); 87.4% (F1-*f*0 x F3-F2, Mel); 86.3% (F1-*f*0 x F3-F2, ERB); 85.8% (F1-*f*0 x F3-F2, LogHz); 85.7% (F1-*f*0 x F3-F2, Koenig-scale); 73.5% (F2-F1 x F3-F2, ERB); 73.3% (F2-F1 x F3-F2, Mel); 72.9% (F2-F1 x F3-F2, LogHz).

[64] The intermediate Hotelling's F values omitted in Table 5 are the following: 1348.286 (F1 x F2-F1, Koenig-scale), 900.005 (F1 x F2-F1 + duration, Koenig-scale); 1330.069 (F1-*f*0 x F2-*f*0, Hz; F1-*f*0 x F2-F1, Hz), 895.572 (F1-*f*0 x F2-*f*0 + duration, Hz; F1-*f*0 x F2-F1 + duration, Hz); 1302.180 (F1-*f*0 x F2-*f*0, Mel), 879.282 (F1-*f*0 x F2-*f*0 + duration, Mel); 1295.823 (F1 x F2, Koenig-scale), 865.417 (F1 x F2 + duration, Koenig-scale); 1280.035 (F1-*f*0 x F2-F1, Koenig-scale), 862.373 (F1-*f*0 x F2-F1 + duration, Koenig-scale); 1267.276 (F1-*f*0 x F2-*f*0, Bark), 856.474 (F1-*f*0 x F2-*f*0 + duration, Bark); 1241.575 (F1-*f*0 x F2-*f*0, ERB), 841.247 (F1-*f*0 x F2-*f*0 + duration, ERB); 1239.296 (F1 x F2-F1, Mel), 827.963 (F1 x F2-F1 + duration, Mel); 1225.504 (F1-*f*0 x F2-*f*0, Koenig-scale), 827.401 (F1-*f*0 x F2-*f*0 + duration, Koenig-scale); 1215.244 (F1-*f*0 x F2, Hz), 819.145 (F1-*f*0 x F2 + duration, Hz); 1215.227 (F1-*f*0 x F2, Mel), 821.643 (F1-*f*0 x F2 + duration, Mel); 1195.215 (F1-*f*0 x F2, Bark), 805.778 (F1-*f*0 x F2 + duration, Bark); 1189.743 (F1-*f*0 x F2, ERB), 807.452 (F1-*f*0 x F2 + duration, ERB); 1189.281 (F1 x F2-F1, Bark), 794.880 (F1 x F2-F1 + duration, Bark); 1159.405 (F1-*f*0 x F2-*f*0, LogHz), 787.875 (F1-*f*0 x F2-*f*0 + duration, LogHz); 1123.893 (F1-*f*0 x F2, Koenig-scale), 759.559 (F1-*f*0 x F2 + duration, Koenig-scale); 1123.473 (F1-*f*0 x F2-F1, Mel), 760.675 (F1-*f*0 x F2-F1 + duration, Mel); 1115.923 (F1-*f*0 x F2, LogHz), 759.798 (F1-*f*0 x F2 + duration, LogHz); 1109.324 (F1 x F2-F1, ERB), 742.151 (F1 x F2-F1 + duration, ERB); 1056.179 (F1-*f*0 x F2-F1, Bark), 716.413 (F1-*f*0 x F2-F1 + duration, Bark); 1056.129 (F1 x F2-F1, LogHz), 706.966 (F1 x F2-F1 + duration, LogHz); 971.156 (F2-F1 x F3-F2, Hz), 647.773 (F2-F1 x F3-F2 + duration, Hz); 917.029 (F1-*f*0 x F2-F1, ERB), 626.339 (F1-*f*0 x F2-F1 + duration, ERB); 904.408 (F2-F1 x F3-F2, Koenig-scale), 603.728 (F2-F1 x F3-F2 + duration, Koenig-scale); 786.652 (F1-*f*0 x F2-F1, LogHz), 541.091 (F1-*f*0 x F2-F1 + duration, LogHz); 726.252 (F2-F1 x F3-F2, Mel), 485.284 (F2-F1 x F3-F2 + duration, Mel); 647.673 (F2-F1 x F3-F2, Bark), 433.101 (F2-F1 x F3-F2 + duration, Bark); 575.080 (F1-*f*0 x F3-F2, ERB), 395.482 (F1-*f*0 x F3-F2 + duration, ERB); 562.824 (F1-*f*0 x F3-F2, Bark), 385.355 (F1-*f*0 x F3-F2 + duration, Bark); 552.516 (F1-*f*0 x F3-F2, Mel), 378.228 (F1-*f*0 x F3-F2 + duration, Mel); 526.321 (F1-*f*0 x F3-F2, LogHz), 364.511 (F1-*f*0 x F3-F2 + duration, LogHz); 505.179 (F2-F1 x F3-F2, ERB), 338.357 (F2-F1 x F3-F2 + duration, ERB).

[65] The intermediate Hotelling's F values not included in Table 6 are the following: 280.259 (F1 x F2, Koenig-scale), 190.453 (F1 x F2 + duration, Koenig-scale); 279.788 (F1 x F2-F1, Koenig-scale), 189.979 (F1 x F2-F1 + duration, Koenig-scale); 266.391 (F1-*f*0 x F2-*f*0, ERB), 180.784 (F1-*f*0 x F2-*f*0 + duration, ERB); 265.378 (F1-*f*0 x F2-*f*0, Mel), 179.521 (F1-*f*0 x F2-f0 + duration, Mel); 265.337 (F1-*f*0 x F2-*f*0, Bark), 179.769 (F1-*f*0 x F2-*f*0 + duration, Bark); 264.993 (F1 x F2-F1, Mel), 180.376 (F1 x F2-F1 + duration, Mel); 260.736 (F1-f0 x F2, ERB), 177.323 (F1-*f*0 x F2 + duration, ERB); 258.480 (F1-*f*0 x F2, Bark), 175.550 (F1-*f*0 x F2 + duration, Bark); 258.021 (F1-f0 x F2, Mel), 174.942 (F1-*f*0 x F2 + duration, Mel); 256.560 (F1-*f*0 x F2-*f*0, Koenig-scale), 173.335 (F1-f0 x F2-f0 + duration, Koenig-scale); 255.727 (F1 x F2-F1, Bark), 174.402 (F1 x F2-F1 + duration, Bark); 254.230 (F1-*f*0 x F2-*f*0, LogHz), 171.756 (F1-*f*0 x F2-*f*0 + duration, LogHz); 253.728 (F1-*f*0 x F2-*f*0, Hz), 170.756 (F1-*f*0 x F2-*f*0 + duration, Hz); 253.728 (F1-*f*0 x F2-F1, Hz), 170.756 (F1-*f*0 x F2-F1 + duration, Hz); 253.176 (F1-*f*0 x F2-F1, Koenig-scale), 170.721 (F1-*f*0 x F2-F1 + duration, Koenig-scale); 247.482 (F2-F1 x F3-F2, Hz), 169.032 (F2-F1 x F3-F2 + duration, Hz); 247.330 (F1-*f*0 x F2, Koenig-scale), 167.346 (F1-*f*0 x F2 + duration, Koenig-scale); 246.707 (F1-*f*0 x F2, LogHz), 166.871 (F1-*f*0 x F2 + duration, LogHz); 245.793 (F1-*f*0 x F2, Hz), 195.765 (F1-*f*0 x F2 + duration, Hz); 241.627 (F1-*f*0 x F2-F1, Mel), 163.628 (F1-*f*0 x F2-F1 + duration, Mel); 238.426 (F2-F1 x F3-F2, Koenig-scale), 164.486 (F2-F1 x F3-F2 + duration, Koenig-scale); 234.383 (F1 x F2-F1, ERB), 160.400 (F1 x F2-F1 + duration,

ERB); 234.188 (F1-*f*0 x F2-F1, Bark), 158.901 (F1-*f*0 x F2-F1 + duration, Bark); 217.658 (F1 x F2-F1, LogHz), 149.285 (F1 x F2-F1 + duration, LogHz); 214.299 (F1-*f*0 x F2-F1, ERB), 146.245 (F1-*f*0 x F2-F1 + duration, ERB); 214.129 (F2-F1 x F3-F2, Mel), 147.629 (F2-F1 x F3-F2 + duration, Mel); 200.525 (F2-F1 x F3-F2, Bark), 138.591 (F2-F1 x F3-F2 + duration, Bark); 187.332 (F1-*f*0 x F2-F1, LogHz), 128.012 (F1-*f*0 x F2-F1 + duration, LogHz); 173.015 (F2-F1 x F3-F2, ERB), 119.680 (F2-F1 x F3-F2 + duration, ERB); 154.111 (F2-F1 x F3-F2, LogHz), 106.649 (F2-F1 x F3-F2 + duration, LogHz); 133.396 (F1-*f*0 x F3-F2, LogHz), 91.149 (F1-*f*0 x F3-F2 + duration, LogHz); 133.005 (F1-*f*0 x F3-F2, ERB), 91.183 (F1-*f*0 x F3-F2 + duration, ERB); 126.239 (F1-*f*0 x F3-F2, Bark), 86.427 (F1-*f*0 x F3-F2 + duration, Bark).

[66] The intermediate Hotelling's F values not included in Table 7 are the following: 488.542 (F1-*f*0 x F2-F1, Koenig-scale), 359.018 (F1-*f*0 x F2-F1 + duration, Koenig-scale); 475.907 (F1-*f*0 x F2-F1, Mel), 344.598 (F1-*f*0 x F2-F1 + duration, Mel); 467.049 (F1 x F2-F1, Koenig-scale), 344.588 (F1 x F2-F1 + duration, Koenig-scale); 453.607 (F1-*f*0 x F2-F1, Bark), 327.008 (F1-*f*0 x F2-F1 + duration, Bark); 449.607 (F1 x F2-F1, Mel), 327.659 (F1 x F2-F1 + duration, Mel); 428.407 (F1-*f*0 x F2-F1, ERB), 303.033 (F1-*f*0 x F2-F1 + duration, ERB); 425.974 (F1 x F2-F1, Bark), 309.105 (F1 x F2-F1 + duration, Bark); 400.222 (F1-*f*0 x F2-F1, LogHz), 278.708 (F1-*f*0 x F2-F1, LogHz); 396.200 (F1 x F2-F1, ERB), 283.180 (F1 x F2-F1 + duration, ERB); 390.296 (F2-F1 x F3-F2, Hz), 284.979 (F2-F1 x F3-F2 + duration, Hz); 373.052 (F1 x F2-F1, LogHz), 263.707 (F1 x F2-F1 + duration, LogHz); 335.746 (F1-*f*0 x F3-F2, Mel), 223.720 (F1-*f*0 x F3-F2 + duration, Mel); 335.166 (F1-f0 x F3-F2, ERB), 223.203 (F1-*f*0 x F3-F2 + duration, ERB); 335.070 (F1-*f*0 x F3-F2, Bark), 223.134 (F1-*f*0 x F3-F2 + duration, Bark); 326.171 (F2-F1 x F3-F2, Koenig-scale), 236.925 (F2-F1 x F3-F2 + duration, Koenig-scale); 325.740 (F1-*f*0 x F3-F2, Koenig-scale), 217.057 (F1-*f*0 x F3-F2 + duration, Koenig-scale); 325.088 (F1-*f*0 x F3-F2, Hz), 217.314 (F1-*f*0 x F3-F2 + duration, Hz); 321.860 (F1-*f*0 x F3-F2, LogHz), 214.284 (F1-*f*0 x F3-F2 + duration, LogHz); 307.591 (F2-F1 x F3-F2, Mel), 223.204 (F2-F1 x F3-F2 + duration, Mel).

Bibliographical References

ADANK Patti (1999), "Acoustic vowel normalisation: Dealing with different sources of variation", in BERGSMA W., M. PALMEN & M. WESTER, eds, *Proceedings of the CLS Opening of the Academic Year 1999-2000*, Nijmegen: 55-77.

ADANK Patti, Vincent J. VAN HEUVEN & Roeland VAN HOUT (1999), "Speaker normalization preserving regional accent differences in vowel quality", *ICPhS99*, San Francisco: 1593-1596.

AINSWORTH William A. (1975), "Intrinsic and extrinsic factors in vowel judgements", in FANT Gunnar & Mark A.A. TATHAM, eds, *Auditory Analysis and Perception of Speech*, London / New-York / San Francisco, Academic Press: 103-113.

BARRY William & Bistra ANDREEVA (2001), "Cross-languague similarities and differences in spontaneous speech patterns", *Journal of the IPA* 31: 51-66.

BERTINETTO Pier Marco, ed. (2001), *AVIP - Archivio di Varietà di Italiano Parlato*, 4 CD-Rom, Pisa, Ufficio Pubblicazioni della Classe di Lettere della Scuola Normale Superiore.

CALAMAI Silvia (2001), "Aspetti qualitativi e quantitativi del vocalismo tonico pisano e livornese", *Rivista Italiana di Dialettologia* 25: 153-207.

CALAMAI Silvia (2002), "La percezione al quadrato in Toscana: pisani e livornesi", in CINI Monica & Riccardo REGIS, eds., Atti del Convegno Internazionale *Che cosa ne pensa oggi Chiaffredo Roux? Percorsi della dialettologia percettiva all'alba del nuovo millennio*, Bardonecchia, 25-27.V.2000, Alessandria, Edizioni dell'Orso: 139-171.

CIONI Lorenzo (2001), "Multi-Speech e ESPS: tecniche di scripting per l'analisi acustica", *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore* 2 (n.s.): 125-137.

COSI Piero, Franco E. FERRERO & K. VAGGES (1995), "Rappresentazioni acustiche e uditive delle vocali italiane", in A. COCCHI, ed., *Atti del XXIII Convegno Nazionale AIA*, Bologna: 151-156.

CROCCO Claudia, Renata SAVY & Franco CUTUGNO, eds. (2002), *API – Archivio del Parlato Italiano*, dvd.

DI BENEDETTO Maria-Gabriella (1989), "Vowel representation: Some observations on temporal and spectral properties of the first formant frequency", *The Journal of the Acoustical Society of America* 86: 55-66.

DI BENEDETTO Maria-Gabriella (1991), "Complex relation between F1 and *F*0 in determining vowel height: Acoustic and perceptual evidence", *Studi Italiani di Linguistica Teorica ed Applicata* 20: 579-603.

DI BENEDETTO Maria-Gabriella (1994), "Acoustic and perceptual evidence of a complex relation between F1 and *F*0 in determining vowel height", *Journal of Phonetics* 22: 205-224.

ESCA (1991), *Proceeding of the ESCA Workshop. Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication*, Barcelona, 30.IX-2.X.1991.

FANT Gunnar (1973), *Speech Sounds and Features*, MIT, Cambridge MA.

FANT Gunnar, Rolf CARLSON & Björn GRANSTRÖM (1974), "The [e]-[ø] ambiguity", *Speech Communication Seminar*: 117-121.

FERRARI DISNER Sandra (1980), "Evaluation of vowel normalization procedures", *The Journal of the Acoustical Society of America* 67: 253-261.

FERRERO Franco E. (1994), "Riflessione sui 'Diagrammi di esistenza delle vocali italiane' dopo 25 anni: evoluzione delle ricerche e prospettive", in FERRERO Franco E. & Emanuela MAGNO CALDOGNETTO, eds, *Le vocali: dati sperimentali, problemi linguistici, applicazioni tecnologiche*. Atti delle III Giornate di Studio del GFS, Padova 19-20.XI.1992: 9-25.

FERRERO Franco E. (1995), "Users' Manual of C:\CSL50\MACROS\USER\FORMANTI.FRA", *Quaderni del CSRF* 14: 385-402.

FUJISAKI Hiroya & Takako KAWASHIMA (1968), "The roles of pitch and higher formants in the perception of vowels", *IEEE Transactions on Audio and Electroacoustics* 16: 73-77.

HEID, Sebastian J.G.G. (1997), "Individual differences between vowel systems of German speakers", *ESCA Eurospeech*, Rhodes, Greece: 759-762.

HERMES Dik J. & J.C. VAN GESTEL (1991), "The frequency scale of speech intonation", *The Journal of the Acoustical Society of America* 90: 97-102.

HILLENBRAND James M. & Robert T. GAYVERT (1993), "Vowel classification based on fundamental frequencies and formant frequencies", *Journal of Speech and Hearing Research* 36: 694-700.

HILLENBRAND James M., Laura A. GETTY, Michael J. CLARK & Kimberlee WHEELER (1995), "Acoustic characteristics of American English vowels", *The Journal of the Acoustical Society of America*, 97: 3099-3111.

HIRAHARA Tatsuya & Hiroaki KATO (1992), "The effect of *F*0 on vowel identification", in TOHKURA Yoh'ichi, Eric VATIKIOTIS-BATESON & Yoshinori SAGISAKA, eds., *Speech Perception, Production and Linguistic Structure*, Ohmsha & IOS Press, Tokyo & Amsterdam: 89-112.

HUANG Caroline B. (1992), "Modelling human vowel identification using aspects of formant trajectory and context", in TOHKURA Yoh'ichi, Eric VATIKIOTIS-BATESON & Yoshinori SAGISAKA, eds., *Speech Perception, Production and Linguistic Structure*, Ohmsha & IOS Press, Tokyo & Amsterdam: 43-61.

GERSTMAN Louis J. (1968), "Classification of self-normalized vowels", *IEEE Trans. Audio Electroacoustic.*, AU-16: 78-80.

KOENIG W. (1949), "A new frequency scale for acoustic measurements", *Bell Laboratories Record* 27: 299-301.

KUWABARA Hisao (1985), "An approach to normalization of coarticulation effects for vowels in connected speech", *The Journal of the Acoustical Society of America* 77: 686-694.

IIVONEN Antti (1994), "A psychoacoustical explanation for the number of major IPA vowels", *Journal of the IPA* 24: 73-90.

LADEFOGED Peter & Donald E. BROADBENT (1957), "Information conveyed by vowels", *The Journal of the Acoustical Society of America* 29: 98-104.

LADEFOGED Peter, Iris KAMENY & William BRACKENRIDGE (1976), "Acoustic Effects of Style of Speech", *The Journal of the Acoustical Society of America* 59: 228-231.

LINDAU Mona (1978), "Vowel Features", *Language* 54: 541-563.
LLISTERRI Joaquim, Dolores POCH-OLIVÉ, eds. (1992), "Special Issue on Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech Communication", *Speech Communication* 11.
LOBANOV B.M. (1971), "Classification of Russian vowels spoken by different speakers", *The Journal of the Acoustical Society of America* 49: 606-608.
MILLER R.L. (1953), "Auditory tests with synthetic vowels", *The Journal of the Acoustical Society of America* 25: 114-121.
NEAREY Terrance M. (1977), *Phonetic feature systems for vowels*, IU Linguistic Club, Bloomington, Indiana.
NEAREY Terrance M. (1989), "Static, dynamic, and relational properties in vowel perception", *The Journal of the Acoustical Society of America* 85: 2088-2113.
POTTER Ralph K. & J.C. STEINBERG (1950), "Toward the specification of speech", *The Journal of the Acoustical Society of America* 22: 807-820.
ROSNER Burt S. & J.B. PICKERING (1994), *Vowel Perception and Production*, Oxford University Press, New York, Oxford.
SALZA Pier Luigi (1991), "La problematica della segmentazione del segnale vocale", in MAGNO CALDOGNETTO Emanuela & Franco FERRERO, eds., *Trattamento del segnale vocale ed elaborazione statistica dei dati*. Atti delle Ie Giornate di Studio del Gruppo di Fonetica Sperimentale (AIA), Padova 3-6.XI.1990, "Collana degli Atti dell'Associazione Italiana di Acustica" 18: 23-48.
SCOTT Brian L. (1976), "Temporal factors in vowel perception", *The Journal of the Acoustical Society of America*, 60: 1354-65.
SIMPSON Adrian P. & Matthias PÄTZOLD, eds. (1996), "Sound patterns of connected speech description, models and explanation", Proceedings of the Symposium held at Kiel University on 14-15 June 1996, *AIPUK* 31.
SYRDAL Ann K. (1985), "Aspects of a model of the auditory representation of American English Vowels", *Speech Communication* 4: 121-135.
SYRDAL Ann K. & H.S. GOPAL (1986), "A perceptual model of vowel recognition based on the auditory representation of American English vowels", *The Journal of the Acoustical Society of America* 79: 1086-1100.
TRAUNMÜLLER Hartmut (1981), "Perceptual dimension of openness in vowels", *The Journal of the Acoustical Society of America* 69: 1465-75.
TRAUNMÜLLER Hartmut (1990), "A note on hidden factors in vowel perception experiments", *The Journal of the Acoustical Society of America* 88: 2015-2019.
UMEDA Noriko & Ryunen TERANISHI (1966), "Phonemic feature and vocal feature: synthesis of speech sounds, using an acoustic model of vocal tract", *J. Acoust. Soc. Japan* 22: 195-203.
VAN SON Rob J.J.H. (1993), "Vowel perception: A closer look at the literature", *IFA Proceedings* 17: 33-64.
WAKITA Hisashi (1977), "Normalization of vowels by vocal tract length and its application to vowel identification", *IEEE Trans. Acoust. Speech Signal Process*, ASSP-25: 183-192.
WEENINK David J.M. (2001), "Vowel normalizations with the TIMIT acoustic

phonetic speech corpus", *Proceedings of the Institute of Phonetic Sciences* 24: 117-123.
ZAHORIAN Stephen A. & Amir J. JAGHARGHI (1993), "Spectral shape versus formants as acoustic correlates for vowels", *The Journal of the Acoustical Society of America* 94: 1966-1982.