Spectrum and timbre of sung vowels Gerrit Bloothooft

SPECTRUM AND TIMBRE OF SUNG VOWELS

VRIJE UNIVERSITEIT TE AMSTERDAM

SPECTRUM AND TIMBRE OF SUNG VOWELS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van
doctor in de wiskunde en natuurwetenschappen
aan de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus dr. P.J.D. Drenth,
hoogleraar in de faculteit der sociale wetenschappen,
in het openbaar te verdedigen
op vrijdag 15 maart 1985 te 15.30 uur
in het hoofdgebouw der universiteit,
De Boelelaan 1105

door

GERRIT BLOOTHOOFT

geboren te Alkmaar

Eerste promotor : Prof. dr. ir. R. Plomp
Tweede promotor: Prof. dr. ir. L. C. W. Pols
Referent : dr. L. W. J. Boves

VOORWOORD

Bij de uitvoering en het op schrift stellen van dit onderzoek zijn velen mij behulpzaam geweest die ik op deze plaats wil bedanken.
Prof.dr.ir. R. Plomp, dit werk en hopelijk ook mijn toekomstig werk draagt uw stempel en daar ben ik blij om. Uw wetenschappelijke originaliteit en uw streven naar kwaliteit zijn mij tot een voorbeeld.
Prof.dr.ir. L.C.W. Pols, uw studies in de zeventiger jaren zijn een basis geweest voor dit onderzoek. Met plezier denk ik aan de discussies met U die onveranderlijk leidden tot een verbetering van mijn werk.
Dr. L.W.J. Boves, U heeft als referent vele kritische en verfrissende opmerkingen gemaakt. Ik hoop dat wij nog vele jaren op deze wijze zullen samenwerken.
Prof.dr. J. Sundberg, I am grateful that I could benefit from your knowledge of the singing voice. You taught me the limitations of my investigations and your remarks improved the quality of my articles.
Voorts wil ik bedanken:
Dr. M.E.H. Schouten, voor de sympathieke wijze waarop U te allen tijde bereid was mijn Engels te corrigeren.
Prof.dr. A. Cohen, voor uw bereidheid mij in mijn nieuwe werkkring de mogelijkheid te bieden om dit proefschrift af te ronden.
Alle collega's en vrienden, voor de plezierige werksfeer, de luisterbereidheid en de hulp bij het oplossen van de onvermijdelijke problemen.
Zonder de enthousiaste en inspirerende medewerking van vocalisten en zangpedagogen was dit onderzoek niet mogelijk geweest. Ik dank
Jan Achterkamp, Marius van Altena, Marianne Blok, Jan Bruin, Cora Canne Meyer, Michael Chance, Alison Deighton, Ina van Duin, Max van Egmond, Maarten Flipse, Palle Fuhr, Frank Hameleers, Paul Hameleers, Rita van Harmelen, Christine Harvey, Maarten Koningsbergen, Bernhard Kruyssen, Francien van Leeuwen, Hein Meens, Lucia Meeuwsen, Marga van Meulenaarsgraf, Jard van Nes, Bep Ogterop, Wout Oosterkamp, Aleid Schokking, Coos Schouten, Marion Smit, Charles van Tassel, Alex Vermeulen, Joke de Vin, Lieuwe Visser, Jacobijn Wallis-Angenent, Marian de Way en Joy Workum.

Dit onderzoek werd gesubsidieerd door de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek via de Stichting Taalwetenschap. Ik dank de collega's in de Werkgemeenschap Fonetiek voor hun blijken van belangstelling.
Tenslotte heeft het schrijven van dit proefschrift van jou, Tineke, veel gevraagd. Ik dank je voor je steun die ik niet had kunnen missen.

Cover design: Eric D'Hondt

CONTENTS

PART I SPECTRAL ANALYSIS OF SUNG VOWELS
(an acoustical study)

1 INTRODUCTION 7

2 MATERIAL AND MEASUREMENTS 12

3 SPECTRAL VARIATION DUE TO DIFFERENCES BETWEEN VOWELS, SINGERS, AND MODES OF SINGING 15

3.1 Introduction 16
3.2 Composition of spectral variance 16
3.3 Spectral variance of subsets of the data 23
3.4 Discussion 25

4 THE EFFECT OF FUNDAMENTAL FREQUENCY ON VOWEL SPECTRA 28

4.1 Introduction 29
4.2 Representation of 1/3-octave spectra 30
4.3 Formant space and spectrum space 33
4.3.1 Matching of vowel configurations 33
4.3.2 A theoretical comparison 35
4.4 Spectrum space and fundamental frequency 39
4.4.1 Vowel configurations 39
4.4.2 Basis vectors 42
4.4.3 Grand-average spectra and singer's formant 44
4.5 Discussion 47

5 SPECTRAL CHARACTERISTICS OF DIFFERENCES BETWEEN SINGERS 52

5.1 Introduction 53
5.2 Results 54
5.2.1 Average spectrum of singers 54

5.2.2 Singer differences for various vowels 58
5.2.3 Singer differences for various modes of singing 60
5.3 Discussion 60
5.3.1 Male voice classification 60
5.3.2 Female voice classification 63

6 SPECTRAL CHARACTERISTICS OF DIFFERENCES BETWEEN MODES OF SINGING 66

6.1 Introduction 67
6.2 Results 68
6.2.1 Average spectrum of modes of singing 68
6.2.2 Differences between modes of singing for various vowels 73
6.3 Discussion 74
6.3.1 The dimension soft-loud 74
6.3.2 The dimension pressed-dark 76
6.3.3 Some perceptual implications 76
6.3.4 Some relations between the factors "modes of singing" and "singers" 77

7 THE OVERALL SOUND-PRESSURE LEVEL OF SUNG VOWELS AND ITS SPECTRAL CORRELATES 78

7.1 Introduction 79
7.2 Variation of overall SPL due to differences between vowels, singers, and modes of singing 80
7.3 Spectral correlates of overall SPL 82
7.4 Discussion 85

8 THE SOUND LEVEL OF THE SINGER'S FORMANT 87

8.1 Introduction 88
8.2 Method 88
8.3 Variation in the sound level of the singer's formant due to differences between vowels, singers, and modes of singing 91
8.4 Discussion 95

PART II THE TIMBRE OF SUNG VOWELS (a psycho-acoustical study)

9 INTRODUCTION 99

10 MATERIAL 102

11 NON-VERBAL LISTENING EXPERIMENTS 104

11.1 Introduction 105
11.2 Procedure 105
11.2.1 Listeners 105
11.2.2 Method 105
11.2.3 INDSCAL analysis 106
11.2.4 Matching of perceptual and spectral vowel configurations 107
11.3 Results and discussion 107

12 VERBAL LISTENING EXPERIMENTS 111

12.1 Introduction 112
12.2 Procedure 113
12.2.1 Listeners 113
12.2.2 Semantic bipolar scales 113
12.2.3 Method 114
12.3 Results 115
12.3.1 Reliability 115
12.3.2 Relations between semantic scales 117
12.3.3 Relations between semantic scales and perceptual space 121
12.3.4 Comparison of vowel configurations derived from non-verbal and verbal listening experiments 125
12.3.5 Spectral correlates of perceptual dimensions of vowel subsets 125
12.3.6 Spectral correlates of perceptual dimensions of song phrases 130
12.4 Discussion 131

PART III

13 RETROSPECTION AND PERSPECTIVES 135

13.1 The spectrum of sung vowels 136
13.2 The timbre of sung vowels 139
13.3 Some implications for singing practice 140

SUMMARY 143

SAMENVATTING 146

APPENDICES 149

A A detailed description of intra-vocalic spectral variation 149
B Representation of modes of singing in the formant space and the spectrum space 161
C Data on overall SPL 164

REFERENCES 165

PART I

SPECTRAL ANALYSIS OF SUNG VOWELS

1 INTRODUCTION

At present there is a growing scientific interest in the singing voice, exemplified by the annual symposia 'Care of the professional voice' held at The Juilliard School in New York and the conferences organized and planned within the framework of the International Decade of Research in Singing between 1981 and 1990. Another milestone is the start of the Journal of Research in Singing by the International Association for Experimental Research in Singing in 1978. Contributions are made from a variety of disciplines: voice teaching, laryngology, speech and voice pathology, and voice science, showing the multi-faceted character of the study of the singing voice. A basic thought in this development is that the gap between the different disciplines, speaking different scientific languages, has to be bridged, which may lead to fruitful discussions and new insights. The position of the present study reflects this interaction of disciplines: although it is basically an acoustical study on vowel spectra sung by professional singers (Part I), the perception of the timbre of sung vowels (by singing teachers, singers, and non-musicians) is investigated too, in order to present the practical relevance of the acoustical factors measured (Part II).

In the questions and problems singers and vocal pedagogues encounter we may distinguish three components. These are the problem of how to achieve a good voice quality (the esthetical component), the problem of how to find adequate verbal expressions for the teaching and discussion of its realization (the verbal component), and the problem of how to keep the vocal apparatus healthy and to avoid disastrous damage to the vocal folds due to a faulty singing technique (the medical component). Finding an adequate description and explanation of acoustic and physiological characteristics of voice production is a prerequisite for solving these problems, at least partly. They will never be solved completely because of, for instance, psychological, subjective, and culturally determined factors.

A better understanding of the production mechanisms underlying good voice quality was aimed at in the pioneering work of, among others, Bartholemew, van den Berg, Hirano, Large, and Sundberg. One fundamental result was that singing can be considered within the general framework of voice production: the acoustical theory of voice production (Fant, 1960) also holds for the singing voice (Sundberg, 1975). Nevertheless, singing differs from speaking in a number of ways because of the better physiological control and the optimal use of acoustical possibilities offered by the components of the vocal apparatus: the subglottal system (respiratory organs), the glottal sound source (vocal folds), and the articulatory system (vocal tract and nasal tract). These enable a singer to realize a voice quality which is quite incomparable with speech. Although the scientific study of speech, voice, and singing have up to now resulted in fundamental knowledge of the separate parts of the vocal organ, many details, especially of the acoustic interactions between these components, are still unknown. It has been suggested by Rothenberg (1982) that effects of acoustical interaction may contribute to a great extent to the concept of voice quality. They may be therefore fundamental to singing.

Whereas a great deal of emphasis has been given to research on human sound production, up to now relatively little attention has been paid to the role of human sound perception in relation to singing. An excellent review of present knowledge on perception of singing has been given by Sundberg (1982). Experimentally verified results in this field mainly focus on vowel intelligibility as a function of fundamental frequency, recognition of vocal registers, perceptual determinants of voice classification, and the effect of vibrato on perceived pitch. In many of these perceptual studies listener subjects turn out to use timbre, a fundamental property of sounds, to distinguish between vowels, registers, or singer voice types. Timbre is used, for instance, as the perceptual quality which allows one to judge whether a phonation has been sung in the falsetto register or not. However, characteristics of the perception of timbre itself have hardly been investigated in singing. Yet the importance of timbre cannot be overestimated: most of the vocabulary in vocal pedagogy for the description of voice quality is related to timbre. Timbre is involved when a voice is called light, dark, pressed, warm, mellow, open, covered, and so on; the list of terms seems unlimited. When we ask, however, for an explicit, objective acoustical definition of these terms only fragmentary data are available. This implies that in discussions on the singing voice in which

these terms are employed, a common opinion about their interpretation is assumed. Often this may be a questionable premise, considering the existing confusion and the opposing points of view in voice pedagogy. Therefore, if we endeavour to obtain a better understanding of each other, a study of the acoustic correlates of the many verbal terms in vocal pedagogy seems very desirable. For each term we may then look for (1) to what extent a common opinion really exists among listeners, (2) whether the term describes a unique phenomenon or whether the term is synonymous with others, and (3) what acoustic aspect of singing the term describes.

Timbre, tone-colour, or "Klangfarbe", has been defined by ASA (1960) as that attribute of auditory sensation in terms of which listeners can judge that sounds having the same pitch and loudness are dissimilar. In this definition timbre discriminates between phonations of any two vowels, any two singers, or any two modes of singing, provided that the phonations are realized with the same pitch and loudness. These differences include a large part of the relevant sound variation in singing. In an effort to present acoustic variables underlying timbre, Schouten (1968) mentioned five factors: (1) tonal or noise-like character, (2) the envelope of the frequency spectrum, (3) the temporal envelope, (4) change in spectral or temporal envelope, and (5) the prefix. Plomp (1970), presenting a review on timbre, left dynamic aspects out of consideration and investigated the timbre of steady-state sounds. For these sounds timbre is determined by the frequency spectrum only, because temporal variables are excluded. Plomp demonstrated the multidimensional character of timbre: more than one perceptual dimension was needed to describe timbre differences between sounds.

Part II of this book is devoted to the study of the perception of the timbre of steady-state sung vowels, its verbal description and its acoustic correlates. The limitation to an investigation of steady-state vowels made it possible to conduct well-defined perceptual experiments, but this experimental paradigm may be rather far-off from perception in real singing performance. It was stressed in a review article by Risset and Wessel (1982) that dynamic factors in timbre contribute to the identification and naturalness of musical instruments. We may assume that this will also be the case for the singing voice. Nevertheless, it was our informal observation that many typical characteristics of sung vowels are also present in their steady-state versions, irrespective of their somewhat unnatural character. Of course, this does not obviate the necessity of an investigation into the dynamic aspects of timbre, but considering the lack of perceptual studies on the timbre of steady-state sung vowels, a limitation to these sounds in the present study seemed justified. In addition, however, one perceptual experiment in which song phrases were used will be reported as an indication of the extent to which results for steady-state vowels have a more general validity.

Experiments by Plomp, Pols, and co-workers have shown that the perceptual dissimilarity in the timbre of steady-state sounds can be physically predicted on the basis of the frequency spectra of the sounds. These spectra are measured with 1/3-octave bandpass filters which bear a relationship to the critical bandwidth in human hearing. This method has been succesfully applied to spoken vowels (Pols et al., 1969) and musical sounds (Plomp, 1970). Although more complex models of the peripheral auditory processing of vowel sounds have been proposed (Karnickaja et al., 1975; Bladon and Lindblom, 1981) the subjective dissimilarity in timbre predicted by these models does not deviate much from results of the 1/3-oct spectrum analysis. In this study we verified the prediction, based on differences in 1/3-oct spectra, for the small timbre differences in one and the same vowel sung by different singers up to a fundamental frequency of 392 Hz (Chapter 11). The success of this verification implies that the timbre of sung vowels may be studied on the basis of their 1/3-oct spectra, so without actually performing perceptual experiments. With this perception-oriented interpretation of 1/3-oct spectra in mind, we investigated in Part I spectral variation which we supposed to be representative of all possible kinds of variation, both in spectrum and in timbre, occurring in professional singing.

It would probably have been logical to begin this book with the results of the perceptual experiments in order to validate a perceptual interpretation of the spectral study to be presented afterwards. The experiments, indeed, were performed in this order. However, the two studies were based on different singing materials; the perception study included only limited spectral variation, whereas the spectral study aimed at presenting an overview of all possible kinds of spectral variation. Because we felt that such an overview would facilitate the interpretation of the results of the perceptual experiments we present the spectral study first.

The investigations of vowel spectra were based on 1/3-oct spectra of 3888 vowel sounds, sung by seven professional male and seven professional female singers. The vowel spectra varied systematically as a consequence of the differences between (1) nine phonemes, (2) typical characteristics of

14 male and female singers with voice classifications from bass to soprano, (3) nine modes of singing, for example a "light," "dark," or "pressed" phonation, or a specific register, and (4) six fundamental frequencies ranging from 98 up to 880 Hz. Chapter 3 deals with the extent to which spectra depend on these four factors. More specifically, the differences between spectra associated with the factors "vowels," "singers," and "modes of singing" are described in detail in Chapters 4, 5, and 6, respectively. Two specific sound levels have been investigated in some more detail: the overall sound-pressure level (Chapter 7) and the sound level of the frequency band of the singer's formant (Chapter 8). The variation in these sound levels due to the factors "vowels," "singers," and "modes of singing" will be given as a function of fundamental frequency.

In contrast to this perception-oriented description of spectra there is the traditional production-oriented approach in which spectra are described in terms of formant frequencies, formant bandwidths, and formant levels. These terms stem from the source-filter model of speech production and are attractive because they help us to understand how vowel sounds are generated. Although the source-filter model also applies to sung vowels (Sundberg, 1975), accurate computation of all variables is difficult for the intermediate and high fundamental frequencies in singing. Pols et al. (1972) showed a close correspondence between a description of spectra of spoken vowels in terms of formant frequencies and on the basis of 1/3-oct spectra. This correspondence is elaborated for sung vowels in Chapter 4. Moreover, 1/3-oct spectra also include spectral variation other than that related to formant frequency; all spectral effects related to voice production, whether these are generated in the glottal sound source, in the vocal tract, or are due to their acoustic interaction, are present in 1/3-oct spectra. In our investigations of 1/3-oct spectra in singing we have not limited ourselves to merely a description and a perceptual interpretation, but we have also tried to interpret the origins of the variation. Balancing between the difficulty of ambiguously interpretable 1/3-oct spectra in terms of voice production and the advantage that spectral variation of all possible origins is available in 1/3-oct spectra, we discuss and outline the most likely interpretation of our results in terms of voice production.

2 MATERIAL AND MEASUREMENTS

A total of 3888 vowels, sung by seven professional male singers (six native speakers of Dutch and one English counter-tenor) and seven professional Dutch female singers were recorded in an anechoic room (microphone distance 0.3 m). Although it took some minutes to familiarize the singers with the acoustics of the room, all singers agreed that the unusual acoustic feedback did not disturb their performance. The classifications of the singers ranged from bass to soprano (Table 2.1). The vowels were sung at nominal fundamental frequencies (F_0) of 98 (G2), 131 (C3), 220 (A3), 392 (G4), 659 (E5), and 880 Hz(A5). A tone at these F_0 values was repeatedly presented during the recording sessions to cue the singer. Depending on the singer's classification , the vowels were sung for a particular subset of these fundamental frequencies. For each fundamental frequency nine Dutch vowels /a/, /ɑ/, /i/, /u/, /ɔ/, /œ/, /y/, /ɛ/ and /e/ were sung in the context /h-vowel-t/ with a duration of 1-2 s (see Table 2.2). The subjects were requested to sing these vowels once in each of nine different modes: neutral, light, dark, free, pressed, soft (pianissimo), loud (fortissimo), straight (without vibrato), and extra vibrato. These terms were adopted from singing pedagogy and were known to all singers. The singers were free in their interpretation of these terms, but subsequent spectral studies did not reveal great differences in interpretation among singers, with, in some cases, the exception of the pressed mode of singing (see Section 6.3.2). The singers confirmed that the nine terms are highly representative of all possible variations in singing vowels. For F_0 = 392 Hz the male singers were also asked to sing in a falsetto voice, if possible.

Table 2.1. Classification of the singers.

male singers		female singers	
1	bass	8	alto
2	bass	9	alto
3	baritone	10	alto
4	baritone	11	mezzo-soprano
5	tenor	12	mezzo-soprano
6	tenor	13	soprano
7	counter-tenor	14	soprano

Table 2.2. List of Dutch words, and approximate English equivalents, used for the recordings of the nine different vowels.

IPA symbol	Dutch word	English word
a	haat	fast
ɑ	hat	father
i	hiet	heat
u	hoet	boot
ɔ	hot	short
œ	hut	the
y	huut	minute (French)
ɛ	het	let
e	heet	face

The counter-tenor could also sing at F_0 = 659 Hz, but these phonations will be left out of consideration in order to avoid mixing up male and female data at that F_0 value. Table 2.3 summarizes the conditions measured.

A computer-controlled analog 1/3-oct filter bank was used for the measurement of the spectra (Pols, 1977). The 1/3-oct filters below 400 Hz were replaced by three 90-Hz wide filters centered at 122, 215, and 307 Hz, in order to adapt the band-pass filters to the critical bandwidths of the human ear (in the present experiment, with fixed fundamental frequencies, this modification was, however, not essential). Since vowels produced by

Table 2.3. Participation of the singers and the number of vowel spectra obtained for each fundamental frequency.

fundamental frequency (Hz)	participating singers	number of vowel spectra
	male singers	
98	1,2,3,4	324
131	1,2,3,4,5,6	486
220	1,2,3,4,5,6,7	567
392 modal	1,2,4,5,6	405
392 falsetto	1,2,7	243
	female singers	
220	8,9,10,11,12,13,14	567
392	8,9,10,11,12,13,14	567
659	8,9,10,11,12,13,14	567
880	12,14	
	total	3888

professional singers usually do not contain much energy beyond 4000 Hz, no band-pass filters with center frequencies higher than 4000 Hz were used. In this way we also avoided problems in the analysis of spectra of which the sound level in the frequency band with a center frequency of 5 kHz dropped below the noise level of the measurement system. For each vowel the 1/3-oct spectrum was measured, in dB, every 10 ms, and normalized for overall sound-pressure level (SPL) to eliminate spectral variation due to overall level differences between spectra. From the perception point of view, this normalization approaches an equalization of the loudness of the vowels. The level-normalized spectra were averaged over a period of 300 ms from the stationary part of the vowel which included at least one vibrato period. No essential information was lost in this averaging procedure (the spectral variance of these 30 10-ms samples was for all fundamental frequencies only 20 dB^2 on the average, which is very small compared with most values in Table 3.1). When the fundamental frequency rises, an increasing number of filter bands do not contain partials; these empty filters were excluded from computation. In this way the total number of frequency bands of interest was 14, 14, 11, 8, 6, and 5 for F_0 = 98, 131, 220, 392, 659, and 880 Hz, respectively.

For the measurement of formant frequencies the power spectrum was used, derived by Fourier analysis from the stationary part of the vowel (FFT, 12 kHz sampling frequency, 256 points, Hamming window). The frequency of a formant was determined from the frequencies and sound levels of the constituting harmonics, using simple location rules and a prior knowledge of the intended vowel (van Nierop et al., 1973).

3 SPECTRAL VARIATION DUE TO DIFFERENCES BETWEEN VOWELS, SINGERS, AND MODES OF SINGING

Average 1/3-oct filter spectra of vowels, sung by seven professional male and seven professional female singers, were measured. The material consisted of nine different vowels, sung at six fundamental frequencies (F_0, ranging from 98 Hz up to 880 Hz). For each vowel the singers were requested to sing in the following nine modes : neutral, light, dark, pressed, free, loud, soft, straight, and extra vibrato. To study the origins of spectral variation, the quantity of spectral variance, based on band filter sound levels, was used. For each fundamental frequency and separately for males and females, portions of the total spectral variance associated with the main effects and the interactions of the factors "vowels," "singers," and "modes of singing," were computed. A considerable decrease in total spectral variance was found when F_0 rose from 98 Hz to 880 Hz, due mostly to the reduced spectral variance between vowels. Above about F_0 = 660 Hz spectral variation was dominated by differences related to singers and modes of singing. Additional analyses revealed that for all F_0 values (1) vowel spectra of the tenor and the soprano singers varied more than those of the bass and the alto singers, (2) there was only a slight dependence of spectral differences between vowels on the mode of singing, and (3) the amount of spectral variation in a vowel, sung by different singers with different modes of singing, was vowel dependent.

Paper with R.Plomp, published in J.Acoust.Soc.Am (1984), 75, 1259-1264.

3.1 Introduction

In the investigations reported in this chapter we were interested in the extent to which vowel spectra depend on the factors "vowels," "singers," "modes of singing," and fundamental frequency. Therefore, we will restrict ourselves to a single overall measure of spectral differences. Differences between 1/3-oct spectra are given by differences in the sound level in each frequency band. The variation of the sound level in a frequency band can be described by its variance. (The variance of n values is defined as the squared differences of these values from their mean value, divided by n.) In analogy to human sound perception, the sound level within each frequency band is supposed to be processed independently from that of other bands. Therefore, as a measure of spectral differences between vowels we sum the variances derived for the different frequency bands and will call this quantity the spectral variance. The magnitude of contributions to spectral variance due to the main effects and the interactions of the factors "vowels" (V), "singers" (S), and "modes of singing" (M) will be reported as a function of fundamental frequency, and separately for male and female singers.

3.2 Composition of spectral variance

The spectral variances associated with the main effects and interactions of all factors are presented in Table 3.1. This table will be explained by discussing, as an example, the results for the male singers at F_0 = 98 Hz. At this fundamental frequency, 4 singers sang 9 different vowels, each in 9 different modes of singing. The sound level in 14 frequency bands determined each 1/3-oct spectrum. The first row of Table 3.1 represents the main effect due to the factor V (vowels). For F_0 = 98 Hz, the number, 622 dB^2, is the spectral variance (variances added over 14 frequency bands) derived from 9 vowel spectra, which were each the average of 36 spectra (4 singers, 9 modes of singing each). This spectral variance is more than half of the total spectral variance, 1172 dB^2, which shows that spectral differences between vowels are by far the most important single source of variance for this F_0. The second row represents the main effect associated with the factor M (modes). In this case the number, 211 dB^2, is the spectral variance derived from 9 spectra for the modes of singing, each being the average of 36 spectra (4 singers, 9 vowels each).

Table 3.1. Spectral variance (dB^2) due to main effects and interactions of the factors vowels (V), modes of singing (M), and singers (S) as a function of fundamental frequency. Degrees of freedom (df) are presented as a function of the number of singers (s), see also Table 2.2. All main effects and interactions were significant beyond the 0.01 level, except those marked with an asterisk. The variance equals the sum of squares divided by the total number of vowels (see Table 2.2). Significance was determined from the F values, derived from appropriate mean squares (sum of squares divided by the degrees of freedom). The factor S was a random variable, the factors V and M were fixed variables. Because no replications of the vowels were available the highest order interaction was used as an error estimate. As a consequence the significance of the interaction SxMxV could not be tested.

		MALES					FEMALES			
F_0 (Hz)		98	131	220	392 mod	392 fals	220	392	659	880
	df									
V	8	622	668	545	187	318	594	398	126	29*
M	8	211	184	186	106	99	115	118	51	79
S	s-1	89	103	103	96	69	122	126	116	58
SxV	(s-1)x8	29	59	68	41	34	82	77	27	26
SxM	(s-1)x8	71	76	68	90	44	51	53	39	18
MxV	64	72	39	39	30	50	30	20	9*	20*
SxMxV	(s-1)x64	78	89	77	52	52	70	58	37	20
Total		1172	1218	1086	602	666	1064	850	405	250

In the same way the spectral variance from 4 average spectra for the singers (each based on 81 vowel-mode combinations), 89 dB^2, was obtained and represented in the third row as the main effect associated with the factor singers (S). Rows 4 to 8 represent the spectral variance due to interactions of factors. The interaction SxV, for example, tells us to what extent the singers realized the vowels in an individual way. The spectral variance due to this interaction, 29 dB^2 for F_0 = 98 Hz, is the difference between the spectral variance in 36 spectra (4 singers, 9 vowels each; each averaged over 9 modes of singing) and the sum of the spectral variance of the main effects due to the factors S and V. The interactions SxM = 71 dB^2 and MxV = 72 dB^2 are explained similarly. The three-way interaction SxMxV represents the extent to which the singers realized the mode of singing of the individual vowels in an individual way. The corresponding spectral variance, 78 dB^2, is the difference between the total spectral variance and the sum of the spectral variances due to the main effects and two-way interactions of all factors.

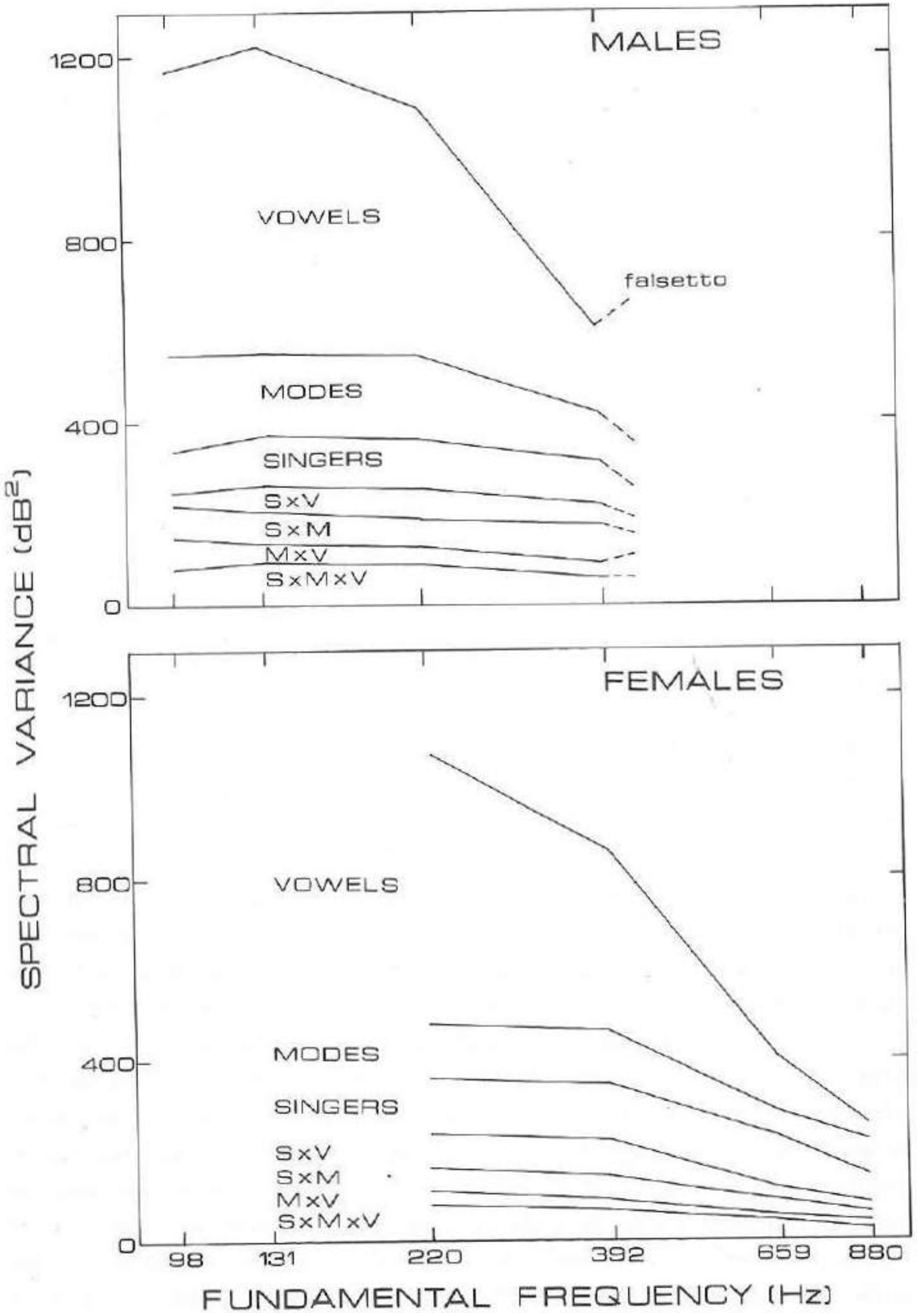

Fig. 3.1. Composition of the spectral variance for the male and female singers as a function of fundamental frequency. Portions of spectral variance are accumulated per F_0 in the indicated order, while the same stages of accumulation for adjacent F_0 values are connected.

A comparison of the components of spectral variance as a function of F_0 is graphically presented in Fig. 3.1. For each F_0 the accumulation of components of spectral variance is given in the following order from below upwards: SxMxV, MxV, SxM, SxV, S, M, and V, where the contribution of a component is added to the preceding sum of contributions. The top value is the total spectral variance. For the sake of clarity, data for male vowels sung in the falsetto register have been shifted a little to the right.

General characteristics of the composition of spectral variance as a function of F_0 are: (1) a considerable decrease of total spectral variance with rising F_0, which is mainly due to the main effect V, and (2) almost all other main effects and interactions are approximately constant up to F_0 = 392 Hz and decrease for higher F_0. This led us to divide components of spectral variance into two categories: the spectral variance between vowels (the main effect V), and the spectral variance within vowels (the sum of the main effects S and M and the interactions SxV, SxM, MxV, and SxMxV). For both categories the variance per frequency band is shown graphically in Fig. 3.2 for all fundamental frequencies. From the production point of view the spectral differences between vowels are of an articulatory nature and are associated with the first two formants. This is reflected in Fig. 3.2 by the dominating contributions of the frequency bands between 500 Hz and 2000 Hz to the spectral variance between vowels. The spectral variance within vowels may have various origins such as the effect of the third and higher formants, the glottal sound source, and the acoustic coupling between glottis and vocal tract. These effects particularly influence the relative amplitudes of the higher harmonics. In agreement with this, Fig. 3.2 shows that the contribution of individual frequency bands to spectral variance within vowels gradually increases with frequency.

For the greater part the reduction of the spectral variance with rising F_0 between vowels in male singing can be explained for F_0 = 98, 131, 220, and 392 Hz (falsetto register) as the effect of the reduction of the number of frequency bands involved. From Fig. 3.2 it can be seen that for these F_0 values the variance in the same frequency band does not vary substantially over F_0. The reduction of spectral variance between vowels therefore merely depends on the loss of variance in the skipped frequency bands. Especially from F_0 = 220 to F_0 = 392 Hz (falsetto register) the reduction is considerable because of the loss of the variance in the important frequency bands with center frequencies of 630 and 1000 Hz. For female singing the same explanation holds for F_0 = 220 and 392 Hz.

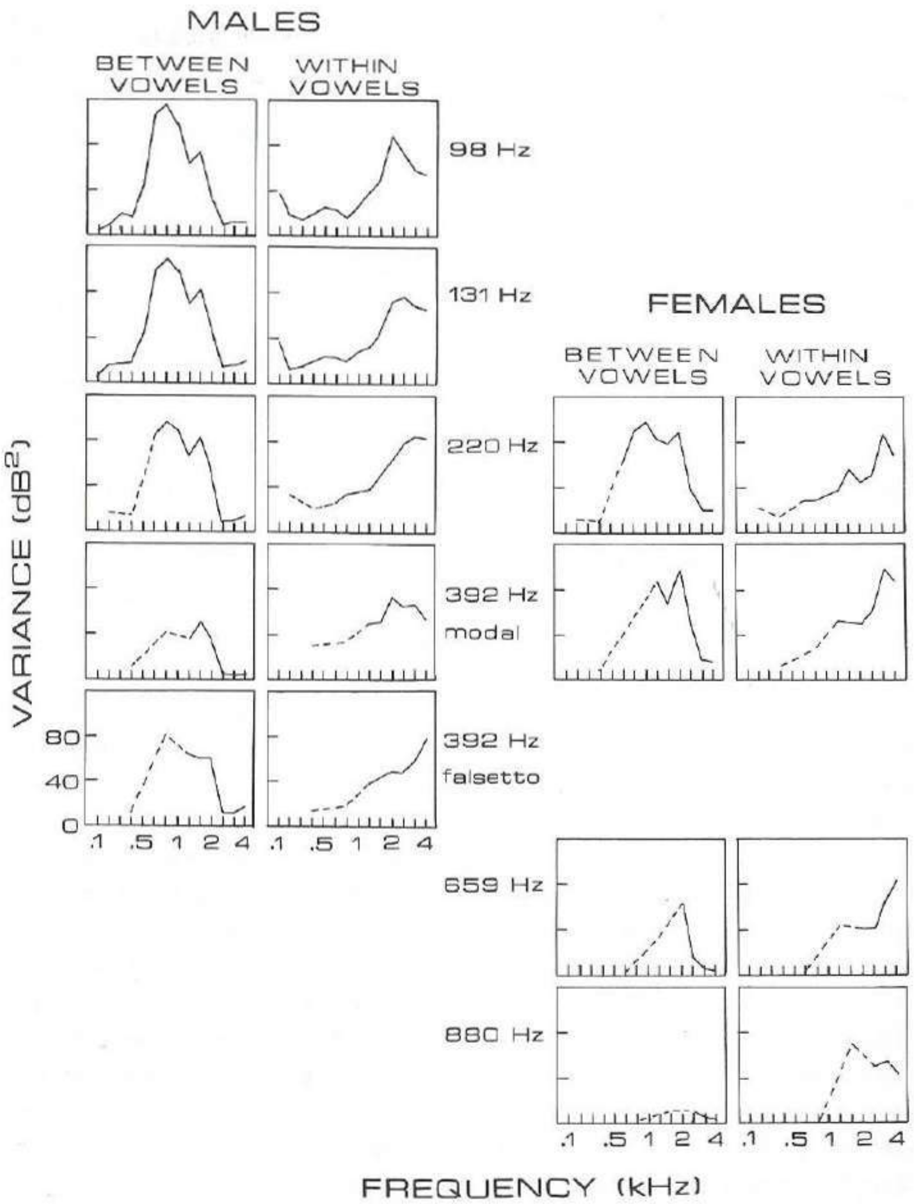

Fig. 3.2. Contribution of frequency bands to the spectral variance between vowels and the spectral variance within vowels as a function of fundamental frequency for males and females. Data for non-adjacent frequency bands are connected by a dashed line.

For both male high-pitched sung vowels (modal register) and female high-pitched sung vowels, however, an additional decrease of the spectral variance between vowels takes place because the variance in the remaining frequency bands is also reduced. For female singers an explanation may be found in a specific way of articulation in which the first formant frequency of high-pitched vowels is tuned to the fundamental frequency to keep a reasonable vocal intensity (Sundberg, 1975). This special articulation will reduce the spectral differences between vowels. For male high-pitched sung vowels it should be noted that the singers experienced many more difficulties in singing vowels in modal register than in singing vowels in falsetto register. Their greater efforts in modal register were probably at the cost of a somewhat reduced articulation, resulting in reduced spectral differences between vowels, while they could maintain a normal articulation in the falsetto register.

The spectral variance within vowels is rather constant up to F_0 = 392 Hz because this variance is not greatly affected by the reduction of the number of frequency bands. This reduction takes place for frequency bands with lower center frequencies which do not contribute substantially to the spectral variance within vowels (see Fig. 3.2). For higher F_0 values the reduction of this variance is in proportion with the number of frequency bands minus one. This proportion results from the property that for these F_0 the fundamental largely determines SPL. Since we normalized the spectra for SPL, the frequency band which contains the fundamental does not contribute to spectral variance.

We may conclude that the dependence of spectral variance on fundamental frequency is highly determined by the number of frequency bands involved. Additionally, it is probable that a reduced or adapted articulation for high-pitched sung vowels resulted for both male and female singers in an extra reduction of the spectral variance between vowels. Because relatively little of the spectral variance within vowels was present in frequency bands with lower center frequencies, this variance was not very dependent on the number of frequency bands up to F_0 = 392 Hz. As a consequence, spectral differences between vowels preponderated for F_0 values less than 392 Hz, while for F_0 values larger than 659 Hz most spectral variation was due to differences between singers, modes of singing, and interactions.

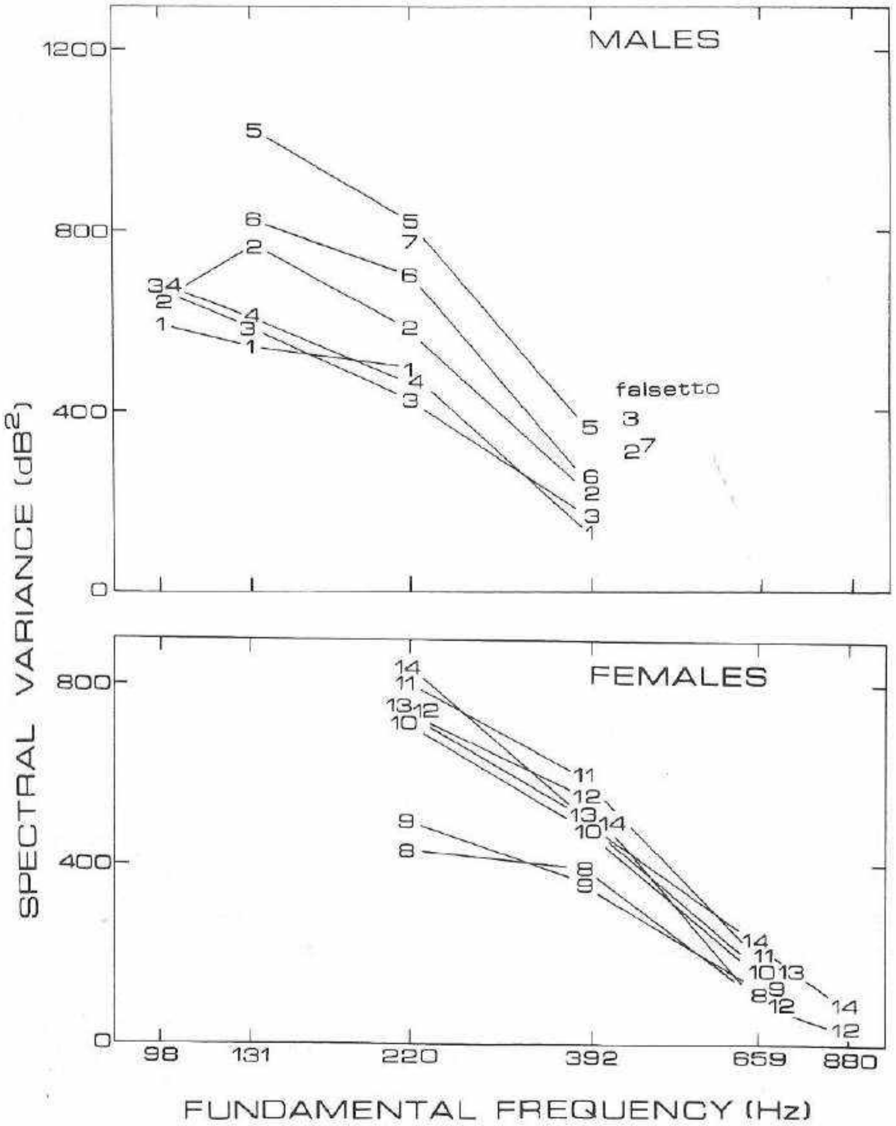

Fig. 3.3. Spectral variance due to differences between vowels, given separately for each individual singer. The numbering of the singers follows Table 2.1.

3.3 Spectral variance of subsets of the data

We can show some more general characteristics of the effects by analyzing subsets of the spectra for each fundamental frequency. First we separately considered the spectra for each fundamental frequency for each singer, and computed the spectral variance between vowels. Fig. 3.3 shows how this variance decreases as a function of F_0 for each singer. The numbering of the singers follows Table 2.1. A systematic relation was found with the classification of the singers: the spectral differences between

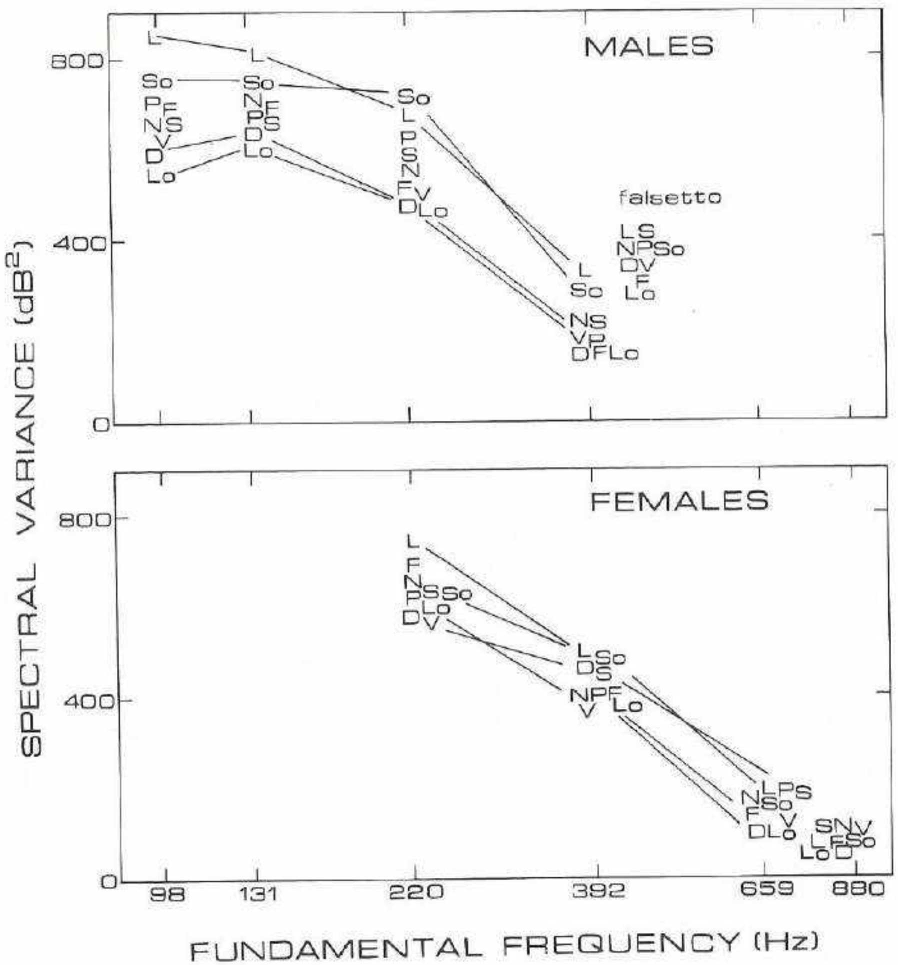

Fig. 3.4. Spectral variance due to differences between vowels, given separately for nine modes of singing: N=neutral, L=light, D=dark, P=pressed, F=free, Lo=loud, So=soft, S=straight, V=extra vibrato.

vowels are larger for the tenor singers than for the bass and baritone singers, and larger for the soprano and mezzo-soprano singers than for the alto singers. This observation holds for most F_0 values except for F_0 = 659 and F_0 = 880 Hz, where the spectral differences between vowels are very small. No relation was found between voice classification and spectral variance due to different modes of singing (not shown).

In the second subset we split up the total number of spectra for each fundamental frequency according to the mode of singing. Fig. 3.4 presents the spectral variance between vowels as a function of F_0 for the nine

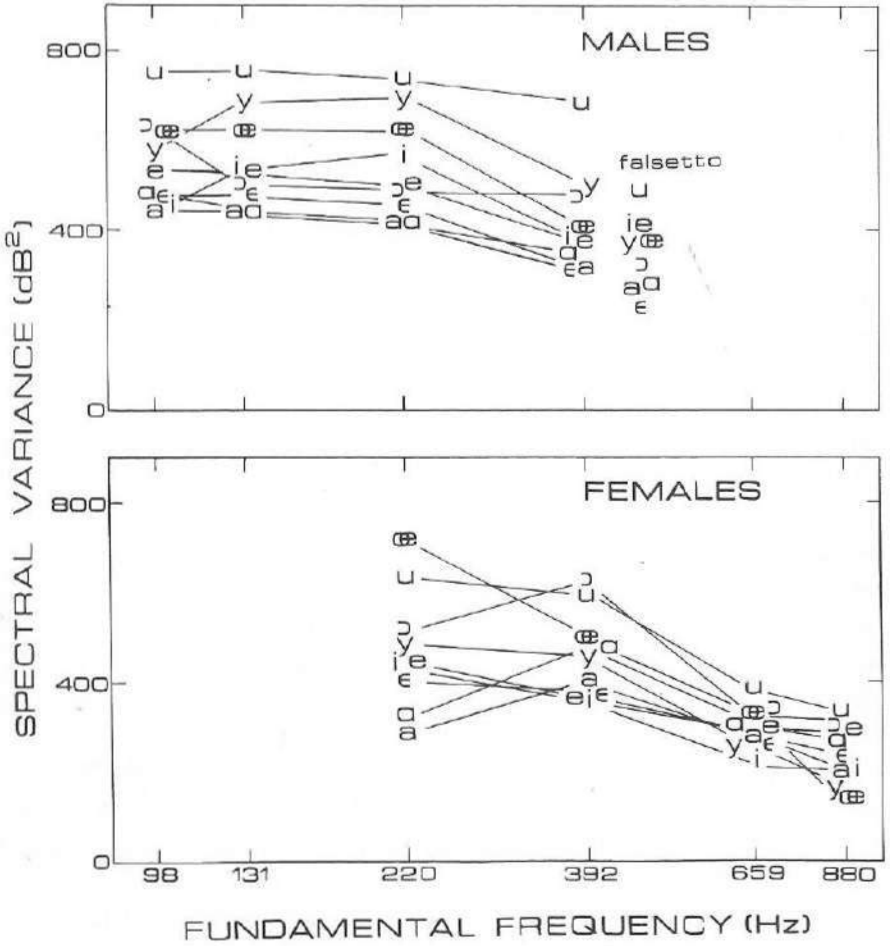

Fig. 3.5. Total spectral variance within a vowel for nine different vowels.

modes of singing. The diagram shows that, especially for the male singers, there is a tendency for the greatest spectral vowel differences to be found for light and soft modes, and the smallest for dark and loud modes. The effect is, however, rather small.

The third subset is the result of separate analyses for each vowel per fundamental frequency. In Fig. 3.5 we give the total spectral variance per vowel as a function of F_0, which includes spectral variation in a vowel due to differences between singers, modes of singing, and their interaction. It is seen that especially for male singers the vowels can be ordered with respect to growing total variance. For F_0 = 131 and 220 Hz we find, ordered from relatively small to large spectral variance: /a/, /ɑ/, /ɛ/, /ɔ/, /e/, /i/, /œ/, /y/, /u/. For female singers a less systematic result was found, probably due to the diminishing differences between vowels with increasing F_0.

3.4 Discussion

Differences in 1/3-oct spectra of stationary vowels are related to subjective dissimilarities in the timbre of the vowels, at least for spoken vowels (Pols et al., 1969; Klein et al., 1970; Plomp, 1970). In Chapter 10 it will be shown that this relation also holds for sung vowels with fundamental frequencies of up to 392 Hz. In this respect the introduction of variance as a measure of spectral variation is important because it is also representative of the spread in subjective dissimilarities. If the mean subjective dissimilarity between vowels decreases, as indicated by a decreasing spectral variance in their 1/3-oct spectra, this will have consequences for the intelligibility of and the confusions between the vowels. The observed F_0 dependence of spectral variance, due to differences between vowels, can therefore explain the well-known decrease in the intelligibility of isolated sung vowels with increasing fundamental frequency (Gottfried, 1980; Chew and Gottfried, 1981; Smith and Scott, 1979,1980). Smith and Scott (1980), however, reported that the intelligibility of isolated high-pitched sung vowels /i/, /I/, /e/ and /æ/ is dependent on the way the vowels have been produced. Vowels sung by a soprano with a raised larynx, which are probably comparable to our description of light or pressed phonation, did not show the drop in intelligibility which was found for operatic sung vowels when F_0 was increased to above 534 Hz. Although we found a slight dependence of vowel-related spectral variance on the mode of singing in favor of light and soft sung vowels, there were no clear indications from a purely spectral point of view supporting the findings of Smith and Scott. Explanations of this discrepancy may be found in (1) a difference between the phonemic distance of vowels and their subjective dissimilarity (Klatt, 1979; Carlson and Granström, 1979); (2) the limited choice among front vowels, (/i/, /I/, /e/, /æ/) of Smith and Scott, which may have favored the raised-larynx sung vowels.

Additionally it should be remarked that the above-mentioned relation between 1/3-oct spectra and subjective dissimilarity has been derived from averaged vowel spectra, comparable with synthetic stationary vowels. Although the effect of temporal variation within our original vowels was shown to be small in terms of spectral variance, the extent of its effect on timbre perception is not precisely known. This limitation should be considered in the interpretation of our spectral data in terms of vowel perception. In relation to the data of Smith and Scott it may be expected that the presence of natural vibrato in their stimuli probably did not play an important role, because Sundberg (1977) did not find an improvement in the intelligibility of synthesized vowels when vibrato was added.

Although there are several studies on the intelligibility of sung vowels, no data seem to be available about the identifiability of singers or speakers as a function of fundamental frequency. Can we distinguish soprano voices more easily than bass voices? General spectral differences between singers are expressed in Table 3.1 by the main effect S; specific realizations of vowels or modes of singing by a singer, which may also contribute to identity, are expressed by the interactions SxV, SxM, and SxMxV. It can be seen in Table 3.1 and Fig. 3.1 that the sum of the spectral variances associated with S, SxV, SxM , and SxMxV is rather constant for F_0 lower than 392 Hz, but decreases in accordance with the decreasing number of frequency bands for higher F_0 values. If we can relate this spectral variance to the identifiability of singers in the same way as we have related spectral variance resulting from the main effect of the factor vowels to the intelligibility of sung vowels, we may conclude that the possibility of identifying a singer's voice from a number of voices would only decrease for high F_0 values. For the same F_0 value, there is no indication that the identifiability of male singers differs from the identifiability of female singers.

We have found a relation between the classification of the singers and spectral variance due to vowel differences (Fig. 3.3). For various F_0 values some tenor and soprano singers showed twice the variance of bass and alto singers, respectively. From the perceptual point of view, this result suggests better intelligibility of vowels sung by the tenor and soprano singers (the higher voice classification) relative to those sung by the bass and the alto singers (the lower voice classification). Because of the small number of singers per voice type, more data will be needed to verify this result.

There was substantial variation among vowels with respect to the total spectral variance per vowel (Fig. 3.5). For males the vowel /u/ always showed the largest variance, for females this was frequently the case, while the vowels /a/, /ɑ/ and /ɛ/ were generally found to have the smallest variance: in some cases about half of that found for /u/. Even with increasing fundamental frequency, and the accompanying decrease in differences between average vowel spectra, this relation was more or less preserved. This result suggests that the tolerance in production as well as the acceptability in perception for the vowel /u/ is probably, over a large frequency range, greater than, that for instance, for the vowel /a/. A more detailed spectral explanation of this result can be found in Chapters 6 and 7.

Our decision to measure and analyze 1/3-oct spectra of sung vowels means that the results were not quantifiable in terms of the parameters of speech production. In spite of this restriction the advantages are: (1) there were no limitations with respect to high fundamental frequencies, (2) both the spectral effects of articulation and glottal variation were included, although not separable, and (3) under the assumption of a relationship between spectral difference and subjective dissimilarity, a perceptual interpretation of the results was possible. In this chapter we restricted ourselves to the investigation of spectral variance. Although this general approach has already allowed us to draw several conclusions, both with respect to the perception and production of sung vowels, a detailed description of the vowel spectra, to be given in the following chapters, will extend the interpretation of the data presented.

4 THE EFFECT OF FUNDAMENTAL FREQUENCY ON VOWEL SPECTRA

For F_o = 98 Hz both an empirical and a theoretical comparison between the representations of vowels in the formant space and the representations in the spectrum space are given. A first-order correspondence between both approaches is demonstrated. A subsequent analysis of the vowel configurations in the spectrum space showed that, apart from differences in the grand-average spectrum, (1) for $F_o \leq$ 220 Hz, vowel configurations were similar for males and females, (2) for F_o = 392 Hz, the variability in dimensions related to F_1 and F_2 was smaller than for lower F_o values, and variability related to F_1 in male falsetto register was greater than in modal register (vowel spectra were similar in male falsetto register and female singing); (3) for F_o = 659 Hz, the vowel configuration had shrunk to clusters of front and back vowels, and (4) for F_o = 880 Hz, vowel differences were only marginal. The relation between the average sound level of the so-called singer's formant and F_o appeared to be vowel dependent. Up to F_o = 392 Hz, the singer's formant was on the average equally prominent for male and female singers, but for higher F_o values its level, relative to overall sound-pressure level, dropped for females. Differences between average spectra of sung and spoken vowels are discussed.

Paper with R.Plomp, accepted for publication in J.Acoust.Soc.Am.

4.1 Introduction

In the previous chapter the spectral variance of vowels sung by seven professional male and seven professional female singers was analyzed. The contributions of the factors "vowels," "singers," "modes of singing", and their interactions were determined as a function of fundamental frequency (F_0). Whereas in that approach differences between vowel spectra were reduced to a single value (spectral variance), in the present chapter we will discuss spectral differences in greater detail by taking the shape of the entire vowel spectrum into account. We will confine ourselves to spectral differences related to the main effect of the factor "vowels", which has been shown to be the most important single source of spectral variance for low F_0 values. In Chapters 5 and 6 spectral characteristics of the factors "singers" and "modes of singing" will be presented. We followed two approaches for the representation of spectral differences: the production-oriented approach in terms of formant frequencies, and the perception-oriented approach in terms of 1/3-oct spectra.

The first approach has the advantage that the formant frequencies are associated with a long tradition in speech analysis, while their articulatory interpretation is well established (e.g. Lindblom and Sundberg, 1971). There are, however, also disadvantages: Spectral effects of the glottal sound source need a separate description which is difficult to obtain, and this approach cannot be applied satisfactorily at high F_0 values, such as may occur in singing, because the wide spacing of the harmonics makes formant frequencies hard to determine.

With the perception-oriented approach, just the reverse is the case. The advantages are that the measurement of 1/3-oct spectra is easy and fast, that source spectrum characteristics are included, and that for the analysis of 1/3-oct spectra narrow spacing of the harmonics is not needed: this approach is applicable to all F_0. One disadvantage is that no theoretical framework exists for the interpretation of the multidimensional representation of 1/3-oct spectra; methods to express spectral differences in a few dimensions are either of a purely statistical nature or based on a correspondence with another representation of the vowels, for instance, one using formant frequencies or perceptual dissimilarities. A second disadvantage is that the varying distribution of harmonics over 1/3-octave bandfilters for varying F_0 necessitates a separate presentation of results for each F_0.

The relevance of this approach for the perception of timbre differences between vowels has been demonstrated for spoken vowels (Pols et al., 1969): The multidimensional vowel configuration based on 1/3-oct spectra matches very well the vowel configuration derived from perceptual dissimilarity judgments. In Chapters 11 and 12 we will show that this close correspondence also holds for sung vowels. Comparable results were reported by Nord and Sventelius (1979), who showed that just-noticeable differences of formant frequencies can be estimated accurately from distance measures based on 1/3-octave spectra. As can be concluded from experiments by Bladon and Lindblom (1981), the advantages of a model still better fitted to the ear's frequency-resolving power are small. They found a correlation coefficient of 0.89 between calculated and judged auditory distances, whereas Plomp (1975), using a prediction based on 1/3-oct spectra, found a value of 0.84. Therefore, the application of a set of 1/3-oct filters is attractive because of its simplicity. Since many readers may not be familiar with the representation of vowel differences by means of 1/3-oct bandfilter analysis, this technique will be explained in Section 4.2.

It has been demonstrated by Pols et al. (1973) and van Nierop et al. (1973) that vowel configurations derived from an analysis of 1/3-oct spectra compare well with formant frequency data for spoken vowels, for both males and females. In Section 4.3 it will be demonstrated for a low value of F_0 that for sung vowels, too, the results of both approaches are very similar. This comparability will be clarified further from a theoretical point of view. On this basis the effect of F_0 on 1/3-oct spectra of vowels will be investigated, and it will be possible to interpret the results in terms of formant frequencies too (Section 4.4). The grand-average vowel spectrum will be investigated for all F_0 values with respect to the level of the so-called singer's formant, the most prominent spectral peak in the range of 2-4 kHz.

4.2 Representation of 1/3-oct spectra

Each vowel spectrum consisted of the sound levels, relative to overall SPL, in 14 frequency bands. The description of these spectra in terms of a "spectrum space" can be illustrated by means of a geometrical model (Plomp et al., 1967). This model consists of a multidimensional Euclidean space in which each dimension corresponds to a frequency band. Each spectrum is represented in this space by a point of which the coordinate values

are the sound levels, relative to overall SPL, for the different frequency bands. Thus, a set of spectra is represented by a cloud of points in the multidimensional space. The center of gravity of the points is associated with the average spectrum. Due to the limited variability of vowel spectra, we may expect the spectra to need far fewer than 14 dimensions for an adequate description. Therefore, we may look for a subspace of the original spectrum space which properly describes the relevant variation between the spectra. Each dimension of such a subspace is a linear combination of all original dimensions. Its direction is given by a basis vector which consists of the direction cosines of the angles between the subspace dimension and all original dimensions. We will call the curve representing the direction cosines as a function of frequency the profile of the basis vector. The coordinate value of a level-normalized 14-dimensional spectrum in a subspace dimension is computed as the inner-product of that spectrum and the basis vector of that dimension.

In terms of this approach, the original spectra can also be interpreted in the following way. Each spectrum consists of the average of all spectra plus a spectral contribution from subspace dimension I, plus a spectral contribution from subspace dimension II, etc. The average spectrum is represented as the origin of the subspace. The spectral contribution of a subspace dimension is equal in shape to the profile of the basis vector for that dimension, while its magnitude is determined by the coordinate value of the spectrum on that dimension.

As was said, we have to look for a subspace which properly describes relevant spectral differences. When another spatial description of the same vowels is available, for instance on the basis of formant frequencies, a subspace can be determined on the basis of the best match between the vowel configurations in both spaces. In that case subspace dimensions approximately describe the spectral effect of formant frequency variation. Without additional data on vowel spectra, subspace dimensions may be determined using statistical properties of the spectra: We may look for a subspace which presents a maximal amount of spectral variation in vowels in a minimum number of dimensions. As a quantitative criterion for this, we can use the percentage of total variance (= sum of squared distances of all points from their center of gravity, divided by the number of points) accounted for, or "explained by", this subspace. For the first subspace dimension that direction is taken that explains as much as possible of the total variance. Subsequently, the second subspace dimension is that direction,

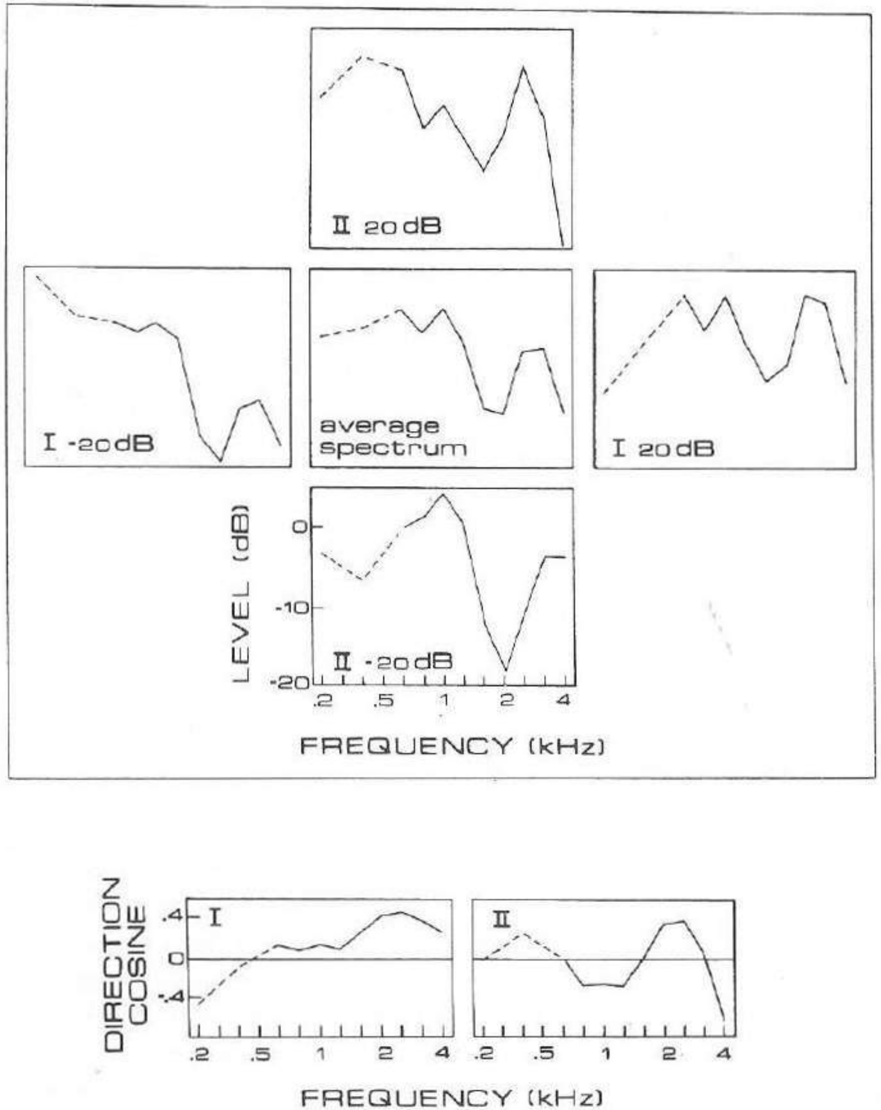

Fig. 4.1. Example of the representation of /a/-like 1/3-oct spectra in two dimensions. The lower panels show the profiles of two basis vectors which determine the spectral variation along dimensions I and II, for which representative spectra are given in the upper panel.

perpendicular to the first one, explaining as much as possible of the remaining variance, etc. Such a subspace can be computed by principal-components analysis (Harman, 1967); the basis vectors of the subspace are the eigenvectors of the variance-covariance matrix, derived from the level-normalized sound levels in all frequency bands of the spectra involved.

An example of how spectral variation can be represented in a spectrum subspace of only two dimensions is given in Fig. 4.1. Five /a/-like spectra are shown: the spectrum in the center is the average /a/ spectrum, in the other four spectra contributions of +20 dB and -20 dB of dimensions I and II are added according to the profiles of the corresponding basis vectors (lower panels). It can be seen that two basis vectors can describe a great variety of spectral variation.

4.3 Formant space and spectrum space

4.3.1 Matching of vowel configurations

As was mentioned above (Section 4.1), sung vowels can be described physically in the production-oriented formant space as well as in the perception-oriented spectrum space. In order to get a better insight into the relation between these approaches, it is of interest to compare for a subset of vowels the configurations in both the formant space and the spectrum space. We used a set of nine vowels, each sung in nine modes of singing by a bass singer at F_0 = 98 Hz. Formant frequencies (F_1, F_2, F_3) were determined for each of these 81 vowel sounds. Since in the present chapter we limited ourselves to spectral differences due to different vowels, data were averaged over the modes of singing. For each vowel the average formant frequencies are plotted in Fig. 4.2 in logarithmic F_1-F_2 and F_1-F_3 planes (filled circles). The axes were chosen logarithmically as a step towards a perception-oriented representation. The configuration of the same nine average vowels in a three-dimensional subspace of the spectrum space is also presented in Fig.4.2 (open circles). This subspace is the one which provides the best match (Schönemann and Carroll, 1970) with the vowel configuration in the logarithmic formant space. The correlation coefficients for the first three dimensions are 0.98, 0.98 and 0.90, respectively. These values are comparable with those obtained by Pols et al. (1973) and van Nierop et al. (1973) for spoken vowels. The percentage of the total spec-

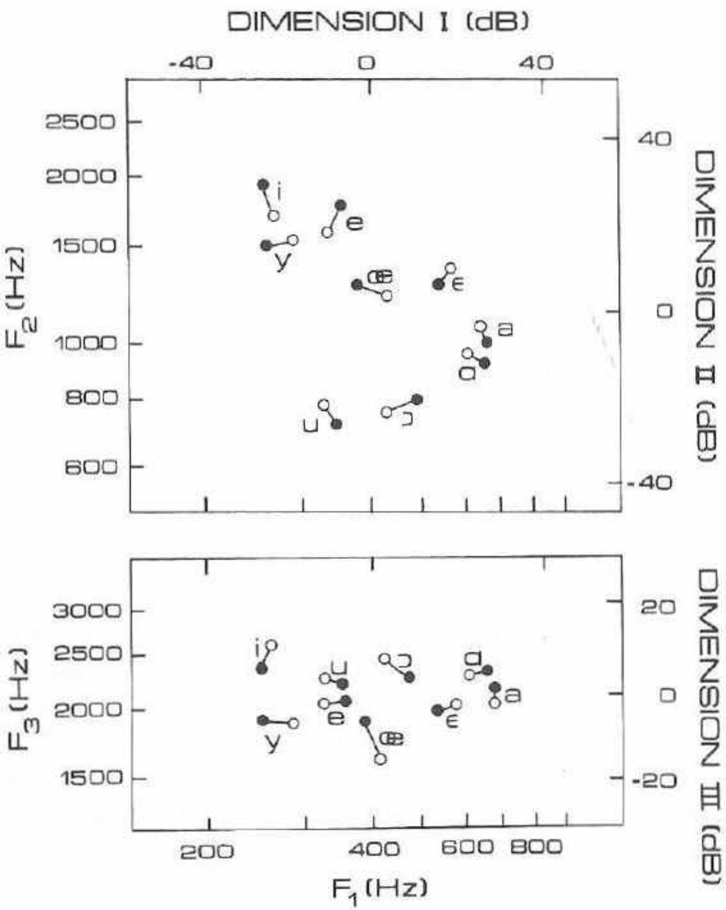

Fig. 4.2. Representation of vowels in dimensions I-II and I-III of the logarithmic formant space (filled circles) and of the best matched spectrum subspace (open circles). Vowel points are each the average of nine versions sung in different modes by a bass singer at F_o = 98 Hz.

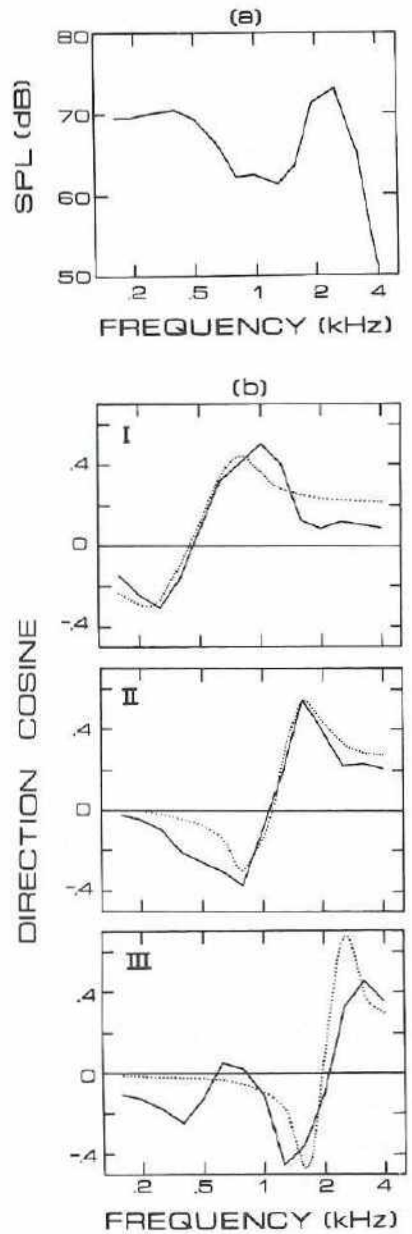

Fig. 4.3. (a) Grand-average spectrum of the vowels presented in Fig. 4.2 (b) Profiles of the basis vectors, related to formant frequency variation on the basis of the matched configurations in Fig. 4.2 (solid lines) and on the basis of the acoustic theory of vowel production (dotted lines).

tral variance explained by each subsequent subspace dimension is 47.7 %(F_1-related), 39.9 %(F_2-related), and 8.7 %(F_3-related), respectively, which shows the importance of the I-II plane. Because of the high percentage of spectral variance included in the three-dimensional subspace, 96.3%, and the high correlations between the first three formant-space and the first three spectrum subspace dimensions, we may conclude that most spectral variation between the average sung vowels is of an articulatory nature, whereas source-spectrum variation, if any, does not play a part.

For the spectrum subspace presented in Fig. 4.2, the grand-average spectrum and the three basis vectors are shown as solid lines in Figs. 4.3a and 4.3b, respectively. Most characteristic of the grand-average spectrum is the high sound level in frequency bands with center frequencies of 2.0 kHz, and 2.5 kHz, showing the average presence of the singer's formant. In the profiles of the basis vectors we find an illustration of the effect of formant frequency variation in terms of 1/3-octave spectra. For example, a positive value along the first spectral dimension implies relatively low levels in frequency bands below 500 Hz and relatively high levels in the 0.5-1.2 kHz region, as can be expected for a high F_1. A negative value along the first spectral dimension implies the opposite: relatively high values in frequency bands below 500 Hz and relatively low values in the 0.5-1.2 kHz region, which is typical of a low F_1. Corresponding relations hold for the second and third dimensions since it can be seen that, apart from a shift in frequency, the three basis vectors have comparable profiles.

4.3.2 A theoretical comparison

Besides the empirical match between formant space and spectrum space and a qualitative interpretation of the meaning of the profiles of the basis vectors of the spectrum space, we can further clarify the relations between both spaces by means of the following theoretical comparison. A number of spectra were computed according to the acoustical theory of speech production (Fant, 1960, p.48). The spectra were computed as transfer functions and were independent of F_0. To investigate how the effect of frequency variation of a single formant is represented in the spectrum space, the first series of spectra had constant values for F_2, F_3, F_4, and F_5 (1300, 2200, 2600, 3000 Hz), but varied in F_1. Typical average values for bass singers were chosen as formant frequencies (see Fig. 4.2; also, Sund-

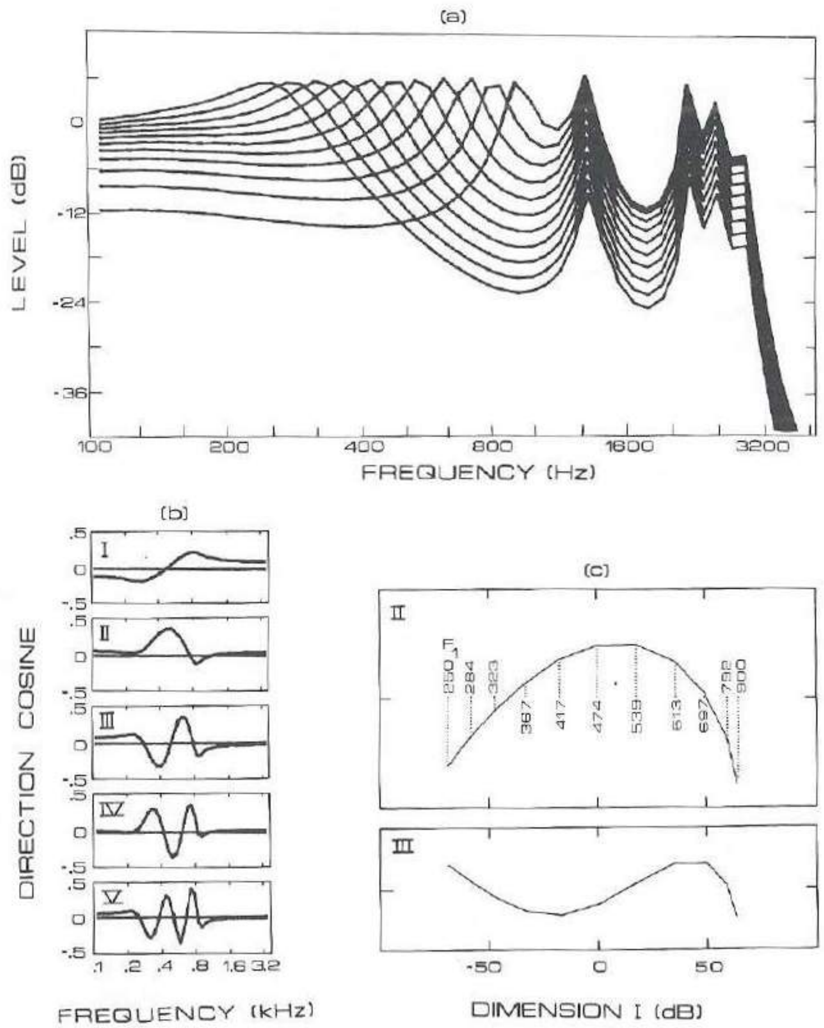

Fig. 4.4. A theoretical approach to the description of formant-frequency variation in the spectrum space.
(a) eleven level-normalized vowel spectra with F_1 varying between 250 and 900 Hz and fixed higher formant frequencies.
(b) Profiles of the first five basis vectors of the space which optimally describes the spectra in (a).
(c) Representation of F_1 variation in the I-II and I-III planes of the spectrum space.

berg, 1975). We used logarithmic axes both for spectrum level and frequency, and varied F_1 in 10 equal logarithmic frequency steps between 250 and 900 Hz. The frequency axis was divided into 50 equal logarithmic intervals from 100 Hz to 4 kHz, and for each value along the frequency axis the relative sound level was computed. The overall SPL of the spectra was equalized, and the resulting spectra are shown in Fig. 4.4a. They can be described, in the same way as 1/3-oct spectra, in a 50-dimensional space according to the subdivision of the frequency axis. Using principal-components analysis, this large number of dimensions could be strongly reduced, while most information was preserved. Fig. 4.4b shows the profiles of the first five basis vectors of the subspace obtained. Percentages of explained variance are 77.5%, 17.5%, 3.0%, 1.3%, and 0.4%, respectively. The large variance in the first dimension shows that, as a first-order approximation, the spectral effect of F_1 variation can be described as a one-dimensional variation in the spectrum space. The variation in F_1 in the spectrum planes I-II and I-III is given in Fig. 4.4c as a curved track. Although the first spectrum dimension gives a good first-order approximation of F_1 variation, more spectral dimensions are needed for a complete description. Notice that the basis vectors I up to V demonstrate a Fourier-like decomposition of the spectral variation, as has also been theoretically proposed by Yilmaz (1967) for speech sounds in general.

It can be inferred from the spectra in Fig. 4.4a that the normalization of overall SPL has a specific influence on the profile of the first basis vector. Overall SPL is determined to a great extent by the first formant. Because of the level-normalization the sound level of F_1 is kept constant, while the low-frequency levels decrease, and high-frequency levels increase, with increasing F_1. This results in a negative weighting of frequency bands with low center frequencies, and a positive weighting of frequency bands with high center frequency, in the basis vector. Furthermore, in so far as higher formants contribute to overall SPL, the choice of their frequency values influences to some extent the profile of the basis vector related to F_1.

What has just been described for the first formant was performed in the same way for two series of vowel spectra, varying in F_2 and F_3, respectively. The average value for F_1 was chosen as 450 Hz and F_2 varied in the second series between 800 and 1800 Hz, while F_3 varied in the third series between 1700 and 2400 Hz. Results were comparable with those obtained for

F_1 and showed a highly one-dimensional representation of variation in F_2 and F_3 in the spectrum space. To compare the theoretical results with empirical data, the profiles of the first basis vectors which were theoretically derived for F_1, F_2, and F_3, are presented as dotted lines in Fig. 4.3b. For optimal comparability the theoretical vectors were scaled to avoid differences due to the different subdivisions of the frequency axis. Although no optimization was aimed at, the correspondence between the theoretical and empirical profiles is remarkably good. Therefore, we may conclude that, to a first-order approximation, sung vowels are presented similarly in the formant space and in the spectrum space. It should be mentioned that the theoretical vectors related to F_1, F_2, and F_3 are not exactly orthogonal, whereas this was necessarily the case for the empirical basis vectors. This may have introduced some additional differences between both sets of vectors, especially in the third dimension, in which the smallest amount of the spectral variance is explained. Furthermore, it is of interest that the theoretical approach did not depend on F_0. This implies that similar shapes of formant-related basis vectors should be expected for all F_0 values, other things being kept constant.

4.4 Spectrum space and fundamental frequency

4.4.1 Vowel configurations

For fundamental frequencies of 98, 131, 220, 392, 659, and 880 Hz, we computed, separately for male and female singers, nine vowel spectra, each of which was the average of spectra over all singers and modes of singing. Frequency bands which did not contain any harmonics were not taken into account; as a result the number of dimensions of the spectrum space decreased from 14 for F_0 = 98 Hz to 5 for F_0 = 880 Hz. To represent the vowel configuration for each F_0 in a spectrum subspace which, firstly, shows optimal spectral differences, and secondly, allows a comparison of vowel configurations over variations in F_0, we applied the following procedure. As a first step we reduced, for each F_0 value, the vowel configuration to five dimensions, using principal-components analysis. These five-dimensional configurations included nearly all the total spectral variance. Then the matching algorithm of Schönemann and Carroll (1970) was applied in an iterative way to rotate these five-dimensional vowel configurations towards maximum congruence. Subsequently, we applied principal-components analysis to all

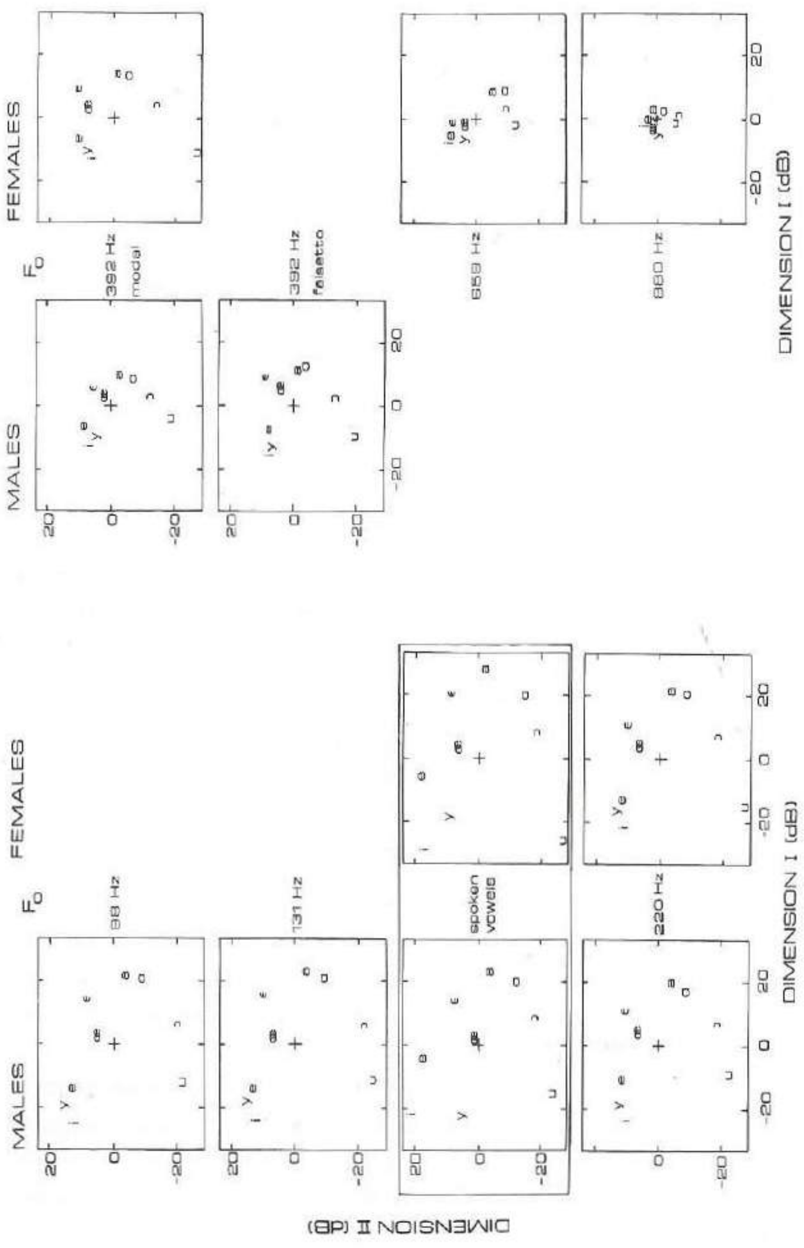

Fig. 4.5. Vowel configurations in the spectrum space for nine vowels, averaged over singers and modes of singing, as a function of F_o. The extra panel represents spoken vowels, adapted from Pols et al. (1973) for males and from van Nierop et al. (1973) for females.

combined vowel points of the matched configurations, and the resulting I-II plane was finally rotated by eye for the best agreement with F_1 and F_2 axes. For each F_o value, the result of these successive operations cumulated in a set of basis vectors for the transformation of the original spectrum space dimensions, which were related to frequency bands, to dimensions related to variation in F_1 and F_2. The resulting vowel configurations are shown in Fig. 4.5. The spectral variance explained in the I-II plane varied between 89.7 % and 98.7 %, indicating that in this plane most of the spectral difference between the vowels was represented. An exception was the vowel configuration for F_o = 880 Hz, in which only 76 % of the total spectral variance was explained. For this F_o value the third dimension, related to F_3, was more important than the first dimension.

In addition, we were able to compare our results with data on spoken vowels. Average spectra of the same Dutch vowels, spoken in the context h-vowel-t, had been obtained earlier by Pols et al. (1973) for 50 males and by van Nierop et al. (1973) for 25 females. The measurement procedure was, apart from minor differences, identical to the present one. For these spoken vowels formant frequencies had also been determined, and the vowels could be represented in the spectrum subspace which allowed the best match with the vowel configuration in the formant space. Correlation coefficients for the first two dimensions were 0.989 and 0.993 for males and 0.973 and 0.991 for females. Results are included in Fig. 4.5 in a separate panel.

A first observation is that, for low F_o, the vowel configurations very much resemble classical configurations in the formant space. For F_o = 98, 131, and 220 Hz, the configurations of sung vowels are almost identical for males and for females (220 Hz only), but there are some distinct differences between sung and spoken vowels: For both males and females the sung front vowels /i/, /y/, and /e/ are clustered, and the difference between the sung vowels /a/ and /ɑ/ is smaller. This is probably the result of a common practice in professional singing, which is to assimilate the qualities of these vowels to each other. There are no further differences between sung and spoken vowels for males, whereas for females the configurations for spoken vowels show more variability in the first dimension (F_1).

With a further increase of F_o we see that the distances between the vowel positions become smaller, in conformity with our findings on the composition of spectral variance (Chapter 3). For the female singers the reduction first largely takes place in the first dimension (F_1), and in the first and second dimensions for the vowels /u/ and /ɔ/. For F_0 = 659 Hz, this results in two clusters: front vowels /i/, /e/, /y/, /œ/, and /ɛ/, and back vowels /u/, /ɔ/, /ɑ/, and /a/. At F_0 = 880 Hz even this difference has almost disappeared, due to a reduction of spectral variation in the dimension related to F_2.

4.4.2 Basis vectors

The profiles of the basis vectors of the spectrum spaces of Fig. 4.5 are shown in Fig. 4.6. As was expected on the basis of the theory, the profiles for values of F_0 up to 392 Hz are almost identical and confirm the profiles of formant-related basis vectors shown in Fig. 4.3b. For F_0 = 392 Hz there is only a small difference between the profiles obtained for phonations in the modal and falsetto registers, indicating that variation in the corresponding spectrum space is similar. Differences between both register types are found in the extent of spectral variation in the first dimension (F_1) and in the grand-average spectrum, as will be seen in the next section. Relative to basis vectors for male singers, for the female singers the weighting of the energy distribution by the basis vectors is shifted slightly to higher frequency ranges, in agreement with the expected average increase of formant frequencies (see panels for F_0 = 220 and 392 Hz). For F_0 = 659 and 880 Hz the basis vector of the second dimension still describes F_2-related spectral variation. However, for these high fundamental frequencies F_1-related spectral variation gradually disappears and is replaced by a weighting of the energy at frequencies higher than 3 kHz.

The profiles for the spoken vowels are based on matching vowel configurations in the spectrum space to those in the formant space. For F_1 the profiles are similar to those of sung vowels. This supports, at least for F_1, the success of our efforts to find, for sung vowels, spectral dimensions related to formant frequencies. Differences in the profile of the second basis vector indicate that for males the average F_2 is lower with sung vowels. For females the profile of the second basis vector shows that with spoken vowels there is not much effect of variation in F_2 on the sound level of frequency bands with a center frequency below 1 kHz, while there is such an effect for sung vowels.

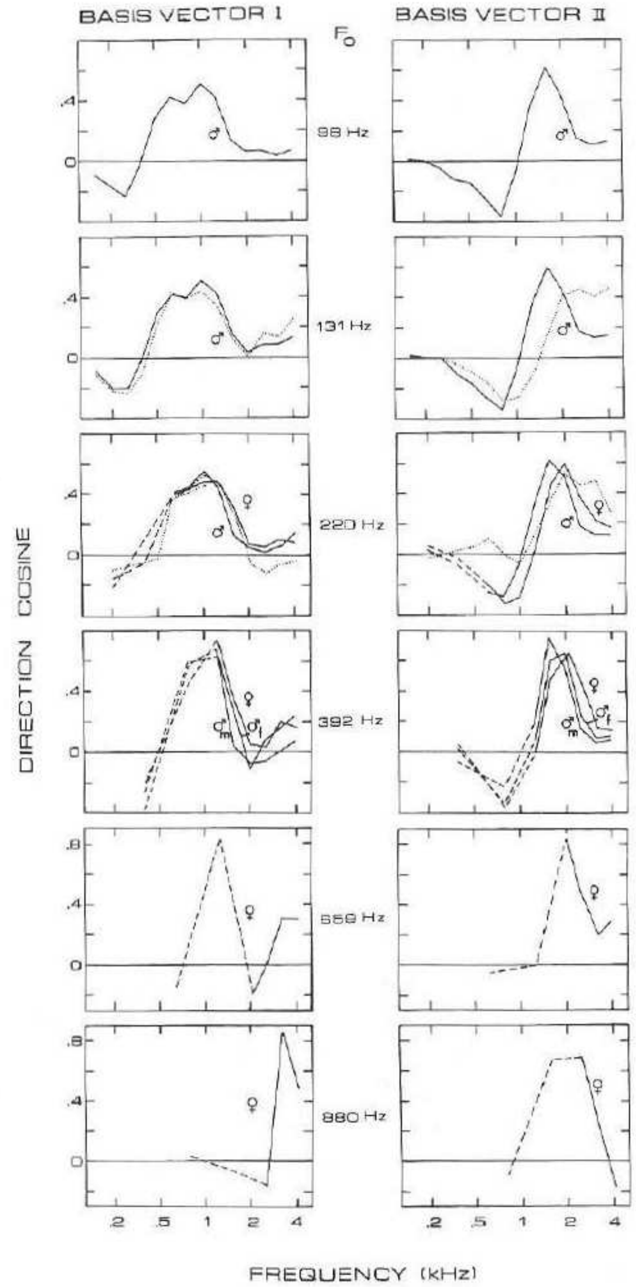

Fig. 4.6. Profiles of the basis vectors of the spectrum spaces of Fig. 4.5. Dashed lines connect data points of non-adjacent frequency bands. Profiles of the basis vectors for spoken vowels (see Fig. 4.5) have been inserted as dotted lines in the panel for F_0 = 131 Hz (males) and F_0 = 220 Hz (females).

4.4.3 Grand-average spectrum and singer's formant

For each F_0 the grand-average spectrum (= origin of the spectrum space) is given in Fig. 4.7 in its original, unnormalized, form. This allows us to study the effect of F_0 on both the shape of the grand-average spectrum and the absolute sound levels of the different frequency bands. Because no information on the average overall SPL of the spoken vowels was available, this level was set arbitrarily at 75 dB SPL. A prominent feature of the grand-average spectra of the sung vowels is the correlation between F_0 and overall SPL. This is shown separately in Fig. 4.8. Overall SPL gradually increases over a range of 16 dB for the male singers when F_0 increases from 98 to 392 Hz, and over a range of 22 dB for the female singers when F_0 increases from 220 to 880 Hz. For the same F_0, overall SPL was on the average 8 dB higher for male than for female singers. For F_0 = 392 Hz, male singers performed on the average at a 9 dB higher overall SPL in the modal register than in the falsetto register. This compares well with differences of between 9 dB and 12 dB reported by Colton (1973) for student singers, singing at the same F_0 in both registers at a comfortable level.

A second striking feature of the grand-average spectra in Fig. 4.7 is the sound level in the frequency bands of the singer's formant. To study properties of the singer's formant in some more detail, we present in Fig. 4.8 for the vowels /a/, /i/, and /u/ the average sound levels in the frequency bands with center frequencies of 2.5 kHz (males) and 3.16 kHz (females). These frequency bands have been chosen because they invariably have the highest sound levels above 2 kHz (see also Fig. 4.7), and because up to now no clear definition of the singer's formant has been given, at least not with respect to its sound level or frequencies of higher formants. Vowels other than /a/, /i/, and /u/ showed intermediate results. For each F_0 the average overall SPL of each separate vowel did not deviate more than 2 dB from the grand-average value and is not included in Fig. 4.8. Recently, the same type of data were reported by Schultz-Coulon et al. (1979) for the vowel /a/ and by Hollien (1983) for the vowel /u/. For comparison, their results are added in Fig. 4.8 as open and filled circles, respectively. Schultz-Coulon et al. measured the relative level of the singer's formant in the 2500-3000 Hz frequency band at three fixed levels of overall SPL for 13 professional male singers and 11 professional female singers. We interpolated their data to match our values of overall SPL. Hollien gave average levels in the 2700-3400 Hz frequency band of 10 professional male

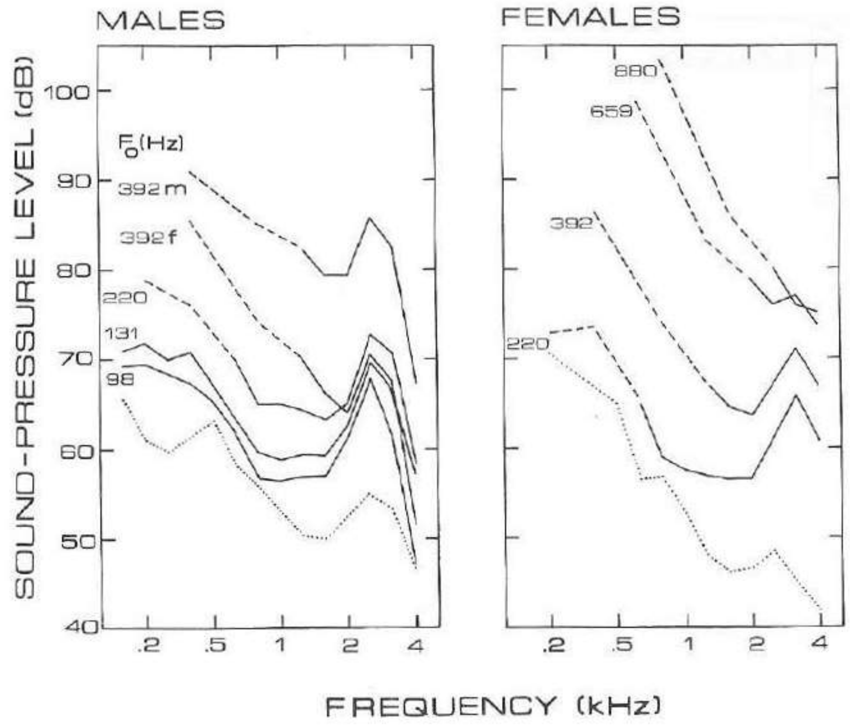

Fig. 4.7. Grand-average spectra of the vowel configurations of Fig. 4.5. Dashed lines connect data of non-adjacent frequency bands. Grand-average spectra of spoken vowels (see Fig. 4.5) are presented as dotted lines.

singers and 10 professional female singers. Since the requested F_0 in his experiments depended on the classification of a singer, we used the average value for both low, medium, and high F_0. In spite of measurement differences, the data of both Schultz-Coulon et al. and Hollien fit in remarkably well with our observations. With male singers the difference between overall SPL and the level of the singer's formant is stable over F_0 in the modal register. In the grand-average spectra (Fig. 4.7) this shows up to some extent as a parallel shift of the sound levels with rising F_0. This was also the case for most separate spectra of the vowels; only the vowel /u/ showed a more rapidly increasing level of the singer's formant. For falsetto phonation the relative level in the frequency band with a center frequency of 2.5 kHz was much reduced compared to modal register.

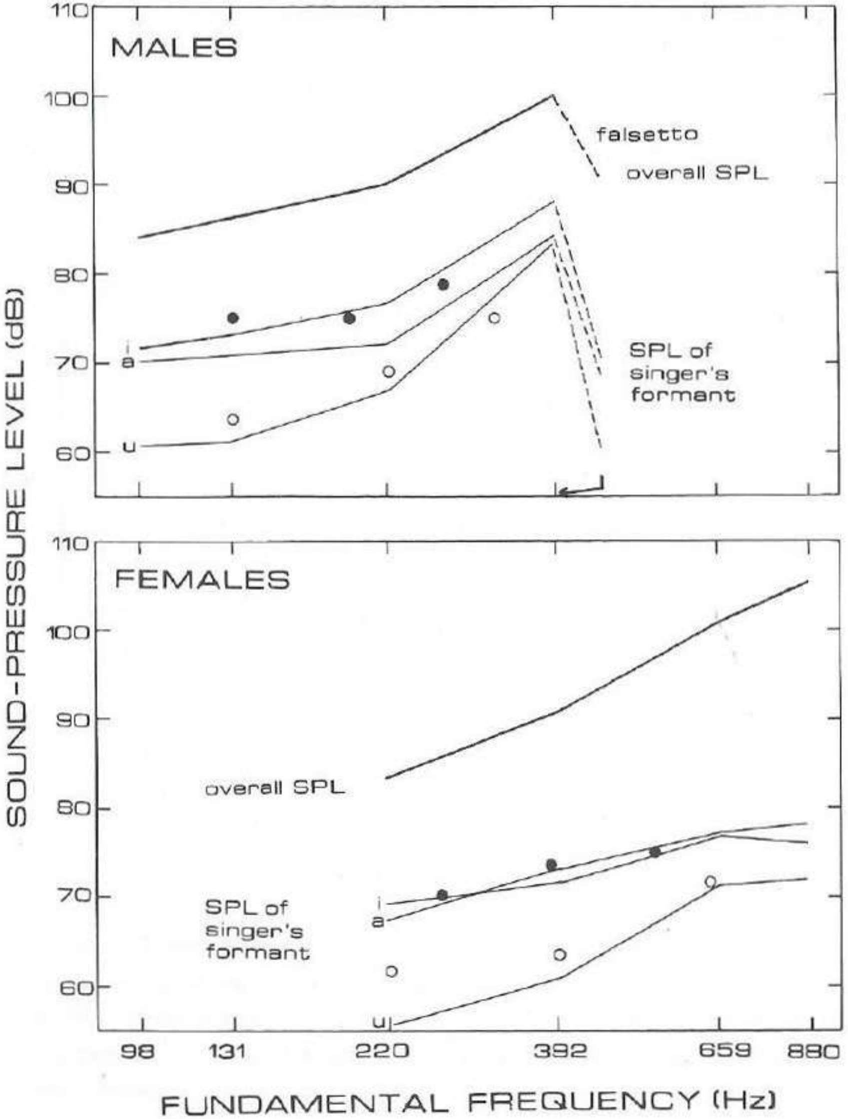

Fig. 4.8. Average overall SPL and average SPL of the singer's formant for the vowels /a/, /i/, and /u/ as a function of F_0 for male and female singers. For the sake of clarity, the data points for the male falsetto register (F_0 = 392 Hz) have been shifted a little to the right. Comparable data on the singer's formant of Schultz-Coulon et al. (1979) for the vowel /a/ (filled circles), and of Hollien (1983) for the vowel /u/ (open circles), have been inserted.

Figure 4.7 shows that, for F_0 = 220 Hz, the grand-average spectra for female singers and male singers were comparable, including the presence of the singer's formant; for F_0 = 392 Hz the female grand-average spectrum most resembles male falsetto phonation. For still higher F_0 the female grand-average spectrum was dominated by the increasing sound level of the fundamental while the level of the frequency band with a center frequency of 3.1 kHz did not increase much.

4.6 Discussion

In this chapter we have concentrated on properties of average vowel spectra. It was shown that a representation of vowels in the formant space and a representation of vowels in a spectrum space give similar results for low F_0, both empirically and theoretically. Because of the difficulty of determining formant frequencies for high F_0, the validity of this correspondence could not be demonstrated for all values of F_0. Nevertheless, the similarity in the profiles of the basis vectors for most F_0 values (Fig. 4.6) suggests that this correspondence holds also for high F_0. Such a similarity should be expected because our theoretical approach in Section 4.3.2 was independent of F_0, and the acoustical theory of vowel production also holds for high F_0 (Sundberg, 1975). In practice, the profiles of the basis vectors were subject to some change with increasing F_0 due to (1) decreasing number of contributing frequency bands, (2) differences in the grand-average spectrum, and (3) decreasing formant-frequency ranges. This variation is not large however, which is also of interest for the application of principal-components based methods for efficient encoding of speech spectra in vocoders (Zahorian and Rothenberg, 1981). Using standard basis vectors, related to variation of formant frequency, these methods may be applicable over a wide F_0 range; in this respect they would compare favorably to many other speech analysis techniques.

If we assume a correspondence between vowel configurations in the formant space and in the spectrum space for all F_0 values, our results, as presented in Fig. 4.5, are in line with those obtained by Sundberg (1975) for a professional soprano singer. Sundberg found (1) decreasing variability in F_1 for F_0 values higher than 262 Hz, since with increasing F_0 the F_1 for low vowels is raised by a wider jaw opening to maintain an acceptable vocal intensity level; (2) decreasing variability in F_2 for F_0 values higher than 400 Hz; and (3) at F_0 = 700 Hz about the same value for both F_1

and F_2 for all vowels. We observed that the first finding probably also applies to a difference between vowels sung and spoken by females: Variability in F_1 is greater in spoken vowels. Sundberg's third finding may have been influenced by the inevitable inaccuracy in the measurement of formant frequencies at such a high F_0 values, or by a specific articulation of the soprano singer. In our analysis we still found considerable variation between front and back vowels for F_0 = 659 Hz (averaged over seven female singers); only for F_0 = 880 Hz did differences between vowels become marginal.

In Part II it will be shown that the distance between vowels in the spectrum space is related to their perceived similarity. A small distance between two different vowels in the spectrum space indicates high perceptual similarity, which may result in confusion of the vowels in an identification task. If we interpret the vowel configurations of Fig. 4.5 in this way, we may expect that for all F_0 values the vowels /a/ and /ɑ/, and the vowels /i/, /y/, and /e/ will be frequently confused. For high F_0 values the distinction between the group of front vowels and the group of back vowels will probably be upheld longest, although vowels within each group will become very similar. Not much perceptual difference between vowels is to be expected for F_0 = 880 Hz. However, these expectations hold for stationary, isolated, vowels only; the beginning and ending of a vowel sound and especially coarticulation have been found to improve intelligibility greatly (e.g., Smith and Scott, 1980).

No formant-frequency data seem to be available for high-pitched male singing in modal and falsetto registers. Our data suggest less variability in F_1 for modal register than for falsetto register (Fig. 4.5). This may be related to problems which arise in high pitched male singing when F_0 comes close to F_1: Sundberg (1981) also observed problems in voice control due to a strong acoustic coupling between glottis and vocal tract. Sundberg suggested that this effect is less prominent in falsetto register or in female singing, due to greater damping caused by a shorter glottal closure. As a result, the F_1 in the male falsetto register and in female singing may have been varied more in the vicinity of F_0 (392 Hz) than in the male modal register. The clearest difference between modal register and falsetto register is to be found in their grand-average spectra (Fig. 4.7), although this difference is in part also caused by the spectral effect of a difference in overall SPL of, on the average, 9 dB. Firstly, the average falsetto spectrum shows a steeper spectrum, in which the frequency band of the fundamental is relatively more important than in the modal register. Secondly, the relative level in the 2.5 kHz frequency band, the singer's formant, is on the average 7 dB lower in the falsetto register than in the modal register. This corresponds to the observation of Vennard (1967, p.89) that only in stronger falsetto phonations a "ring" can be heard. Furthermore, there is a remarkable similarity between results for male falsetto register and female singing at F_0 = 392 Hz, with respect to vowel configuration (Fig. 4.5), profiles of the basis vectors (Fig. 4.6), grand-average spectrum (Fig. 4.7), and overall SPL (Fig. 4.8). Because of the perceptual relevance of the spectrum-space approach, this suggests that a counter-tenor may often be difficult to distinguish from an alto singer. More details of singer-specific spectral characteristics will be presented in Chapter 5.

The grand-average spectrum was generally dominated by variation in overall SPL and by the sound level of the singer's formant as a function of F_0 (Fig. 4.7). Schultz-Coulon et al. (1979) found for the vowel /a/ that, with professional male and female singers, the sound level of the singer's formant relative to overall SPL (1) increases with increasing overall SPL, F_0 being kept constant, and (2) decreases with increasing F_0, overall SPL being kept constant. Most likely, both effects can be attributed to variation in the spectrum of the glottal sound source (Sundberg, 1973; Gauffin and Sundberg, 1980). Since, on the average, overall SPL increased with increasing F_0, the relative level of the singer's formant in our grand-average spectra is governed by the sum of these two opposing mechanisms. For the male singers, in the modal register, both mechanisms were, on the average, approximately in balance for the vowels /a/ and /i/, resulting in a stable difference between overall SPL and the sound level of the singer's formant. For the vowel /u/ the sound level of the singer's formant increased more rapidly than overall SPL, indicating that for this vowel the effect of increasing overall SPL prevailed over the effect of increasing F_0. For the female singers the increasing difference between the sound level of the singer's formant and overall SPL with increasing F_0 is substantial (Fig. 4.8). This indicates that, with respect to the sound level of the singer's formant, for female singing the effect of increasing F_0 dominates over the effect of increasing overall SPL. For F_0 = 220 Hz the shape of the female grand-average spectrum is comparable to the one for males, including the singer's formant (see Fig. 4.7). For F_0 = 392 Hz, however, the female grand-average spectrum compares well with the male grand-average falsetto spectrum. This result supports the observation of

Barthelemew (1934) that the singer's formant is present in female singing (especially in good contraltos), but drops out when F_0 rises above some point. This point probably lies somewhere between 220 Hz and 659 Hz and may result from the spectral effect of an articulation which does not fulfill the requirements of the singer's formant (Sundberg, 1974) or from a steeper source spectrum.

We compared average spectra of sung vowels with those of spoken vowels, from studies by Pols and co-workers. For bass singers Sundberg (1970) found, relative to spoken vowels, that in sung vowels (1) F_2 is lowered in front vowels, (2) F_3 is raised in back vowels and lowered in the other vowels, and (3) F_3 up to F_5 are concentrated for all vowels. The second and the third finding result in the singer's formant. These results were interpreted articulatorily by Sundberg (1970) as the effects of a lowered larynx for all sung vowels, and in addition of an altered tongue shape for back vowels and an increased lip protrusion for sung front vowels. In addition, Sundberg (1974) found that, theoretically, a high amplitude of the singer's formant can be generated by an area mismatch between the pharynx and the entrance of the larynx tube, which is probably promoted by a lowering of the larynx, but not necessarily so. Our vowel configurations (Fig. 4.5) and the profiles of the second basis vector (Fig. 4.6) agree with a lowered F_2 for front vowels, while the grand-average spectrum (Fig. 4.7) shows the average presence of the singer's formant. Because our data were obtained as the averages over seven male singers with a classification from bass to tenor, the results of Sundberg (1970) seem to apply generally to male voices.

We were able to extend the comparison between sung and spoken vowels to female subjects. With respect to F_2 and the singer's formant, the same differences between spoken and sung vowels were found for male and female singers. In addition, it was found that the variation in the F_1-related dimension was larger in female spoken vowels than in sung vowels, and also larger than in male spoken and sung vowels. An important reason for this is the finding of van Nierop et al. (1973) that for the spoken Dutch vowels /i/, /y/, and /u/, F_1 is lower for females than for males, while all other vowels have higher F_1 values for females. Our results indicate that this divergence of F_1 is not present in sung vowels. In general it can be observed that the configurations for male and female sung vowels are very similar, indicating that, apart from the male-female difference in the grand-average spectrum, the same relative differences between vowels are

produced by both sexes.

It is of interest that the difference in sound level between spoken and sung vowels in the frequency bands with high center frequencies is even more pronounced for females than for males (Fig. 4.7). Apart from articulatory differences between spoken and sung vowels as the explanation of the singer's formant (Sundberg, 1974), we probably have to take into account a considerable difference in average overall SPL between spoken and sung vowels, too. No data for the spoken vowels were available. A lower overall SPL in spoken vowels reduces the relative level in the higher frequency bands, and females possibly spoke at the lowest average overall SPL. In this respect it should be mentioned that for a low overall SPL of 75 dB (microphone distance 0.3 m) Schultz-Coulon et al. (1979) did not find significant differences in the relative level of the singer's formant for the vowel /a/ between untrained subjects and professional singers. For the vowel /u/, Hollien (1983) does report differences, although, unfortunately, low vocal intensity was not precisely defined. We refer to Chapter 8, in which the relationship between the level of the singer's formant, F_0, overall SPL, voice classification, and mode of singing will be presented for all nine vowels.

5 SPECTRAL CHARACTERISTICS OF DIFFERENCES BETWEEN SINGERS

The main spectral differences between singers could be described in two dimensions. The first dimension mainly described differences among male singers, the second dimension those among female singers. This suggested a different origin of spectral differences for both sexes, in which the interindividual differences between males and the average difference between males and females have a morphological basis and the interindividual differences between females have a glottal basis.

Part of a paper with R.Plomp, submitted to J.Acoust.Soc.Am.

5.1 Introduction

Most acoustical studies on spectral properties characterizing individual voices concentrate on the average male-female differences in formant frequencies of spoken vowels (for a review, see Fant, 1975). Differences in average formant frequency have been associated with morphological differences between the male and female vocal tracts: a greater pharyngeal length and more strongly developed laryngeal cavities for males. More specifically, Cleveland (1977) showed that there is, in professional male singers, a relation between the average value of the first four formant frequencies of the Swedish vowels /i/, /e/, /ɑ/, /o/, and /u/, and a jury evaluation of voice timbre type. He suggested that the same morphological differences between males and females may distinguish between bass and tenor singers too.

This morphological basis for inter-individual timbre differences between singers was investigated by Dmitriev and Kiselev (1979) for 20 professional singers with voice classifications from bass to high soprano. By means of the X-ray method, they measured the length of the vocal tract, from the lips to the vocal folds, and correlated this length with the frequencies of spectral peaks in the long-term average spectrum in the range of 450-800 Hz and 2300-3500 Hz. A high correlation was found: the longer the vocal tract, the lower the frequencies of both spectral peaks. Each voice was characterized by a typical length of the vocal tract; the ranges in length for the voice types showed hardly any overlap except for mezzo-soprano and soprano singers, who occupied identical ranges.

By studying the contrasts between spectra of vowels sung by alto and tenor singers, Ågren and Sundberg (1978) found that both the amplitude of the fundamental of the source spectrum and the frequencies of the third and fourth formants may differentiate between these two voice types. This result indicates that not only morphological differences but also characteristics of the glottal waveform, which, among other things, influence the amplitude of the fundamental, may distinguish between voice types. For the speaking voice, Monsen and Engebretson (1977) found a steeper glottal spectrum for females than for males, which could be explained, however, by the higher fundamental frequency of the female voice. In the present study, characteristic spectral differences between individual singers, males and females, were investigated over a wide range of fundamental frequencies. Differences between modal register and falsetto register for some male singers will be also given.

5.2 Results

5.2.1 Average spectra of singers

We studied spectral differences between singers for each F_0 value separately. The main differences between singers can be investigated using the differences in the average spectra of the singers. Each of these spectra was the average of nine vowels sung in nine modes of singing (a total of 81 vowels). Since in Chapter 3 we showed that the spectral variance due to differences between singers did not vary much over F_0, we will present results for F_0 = 220 and 392 Hz only. By combining male and female singers we obtained 14 average spectra for F_0 = 220 Hz and 15 average spectra for F_0 = 392 Hz. The latter subset included for some male singers both the average modal spectrum and the average falsetto spectrum. By means of principal components analysis, these spectra could be described very well in a two-dimensional space which explained 76.0 % (F_0 = 220 Hz) and 83.6 % (F_0 = 392 Hz) of total spectral variance. We rotated the first two dimensions of the principal-components solution in such a way that the main differences among male singers and among female singers came out in dimension I and II, respectively. The resulting configurations of points for the singers are shown in Figs. 5.1 and 5.2 (panel a) for F_0 = 220 and 392 Hz, respectively, together with the grand-average spectrum (panel b), the profiles of the basis vectors of the dimensions presented (panels c), the distribution of variance over the various frequency bands (panel d), and typical examples of the average spectra of some singers (panels e). To facilitate the interpretation of the profiles of the basis vectors, they are presented with a polarity corresponding to positive coordinate values in the lower left corner of the I-II plane.

The point configurations indicate that we need at least two dimensions to describe spectral differences between male and female singers. For F_0 = 220 Hz the singer points form approximately a triangle with bass, alto, and soprano singers at the angles. The left part of the triangle represents the male singers who have their major variation, from bass to (counter)-tenor voice classifications, along dimension I (except tenor singer 6). The right part of the triangle is occupied by female singers who all have a lower coordinate value in dimension I than male singers. Among themselves, however, female singers vary in the second dimension from soprano to alto voice classifications.

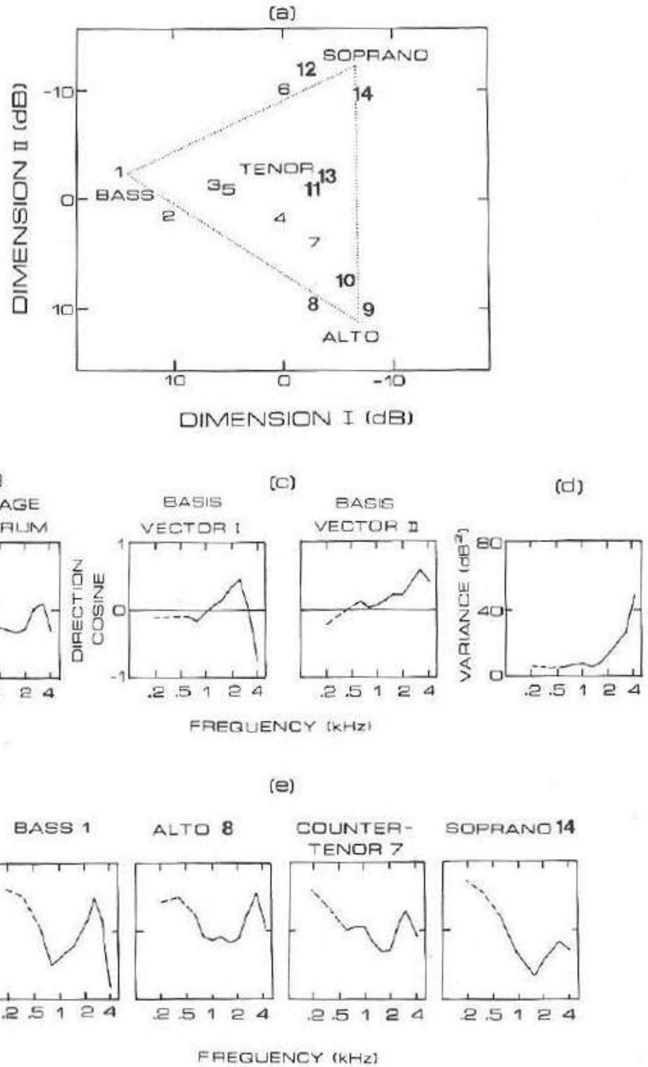

Fig. 5.1. Representation of spectral differences between singers at F_0 = 220 Hz. Light numbers indicate male singers (1-2, bass; 3-4, baritone; 5-6, tenor; 7, counter-tenor), heavy numbers indicate female singers (8-10, alto; 11-12, mezzo-soprano; 13-14, soprano).
(a) Configuration of points of singers in a two-dimensional spectrum subspace.
(b) Grand-average spectrum.
(c) Profiles of the basis vectors of the dimensions presented.
(d) Distribution of spectral variance due to the factor "singers" over frequency bands.
(e) Typical average spectra of four singers.

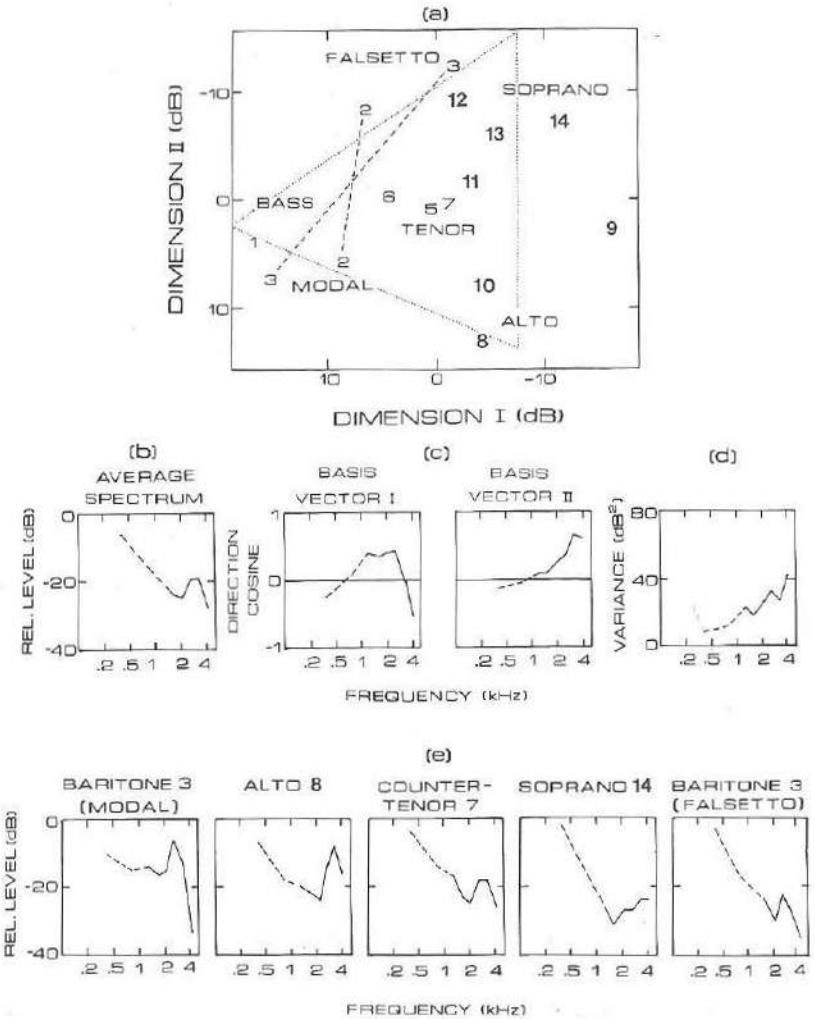

Fig. 5.2. Representation of spectral differences between singers at F_0 = 392 Hz. See Fig. 5.1 for a further explanation of the panels.

The same type of singer point configuration is obtained for F_0 = 392 Hz, although F_0 appears to influence spectral differences between singers to some extent. In this connection it should be realized that F_0 = 220 Hz is near the bottom of the singing range for female singers, whereas F_0 = 392 Hz is near the top of the singing range for male singers. Comparison of the configurations for F_0 = 220 and 392 Hz shows up that di-

mension I presents the more stable differences among male singers and dimension II those among female singers. It is of interest that the average falsetto spectra in the subset for F_0 = 392 Hz are characterized by a low coordinate value for dimension II and resemble average soprano spectra.

The spectral interpretation of the dimensions of Figs. 5.1 and 5.2 can be inferred from the grand-average spectrum (panel b) and the profiles of the basis vectors (panels c). A first observation for both subsets is that most spectral variation between singers is to be found in the higher frequency bands. The distribution of the variance (panel d) shows that more than half of the total spectral variance due to singer differences is found in the frequency bands with center frequencies of 2.5, 3.16, and 4.0 kHz combined. For F_0 = 220 Hz, the basis vector of dimension I, which described the main average male-female and bass-tenor differences, is sensitive to the frequency position of the spectral peak around 2.5 kHz (also known as singer's formant). A positive coordinate value in this dimension, such as in bass singers, characterizes a spectrum with relatively high sound levels in the frequency bands with center frequencies of 2 or 2.5 kHz and a relatively low sound level in the frequency band with a center frequency of 4 kHz. A negative coordinate value, such as in tenor and female singers, implies the reversed spectral properties. The typical spectra in panel e illustrate these properties: the high frequency peak is found for bass singer 1 in the 2.5 kHz frequency band while the sound level in the 4 kHz band is low; for the singers 7, 8, and 14, with about equal coordinate values along dimension I, the peak is found in the 3.1 kHz frequency band. It can be noted that with respect to the high-frequency spectral peak the same results were found for the grand-average spectrum of male and female singers (Fig. 4.7).

The profile of the basis vector of the second dimension, describing the main alto-soprano differences, shows spectral-slope like weigthing properties. A positive coordinate value along this dimension, such as in alto singers, implies a relatively low sound level of the fundamental and relatively high sound levels in the frequency bands with center frequencies beyond 1.6 kHz. The spectral effect of this dimension is illustrated in the average spectrum of alto singer 8 and soprano singer 14 in panel e.

For F_0= 392 Hz (Fig. 5.2) approximately the same spectral interpretation can be given to both dimensions as for F_0= 220 Hz. Panel e presents in addition the average spectrum for the modal register and the falsetto register for baritone singer 3. As a result of the great reduction of the sound levels in the higher frequency bands for the falsetto register, the average spectrum of this singer compares well with the average spectrum of soprano singer 14. For counter-tenor 7 the average spectrum compares better to those of mezzo-soprano singers.

A representation of the differences between singers at other fundamental frequencies than 220 and 392 Hz revealed essentially the same results. The main differences between male singers for F_0 = 98 and 131 Hz resided in the same spectral variation as the one described by dimension I, whereas spectral differences for females at F_0 = 659 Hz compared well with dimension II of Figs. 5.1 and 5.2.

5.2.2 Singer differences for various vowels

It may be asked whether the present findings for singer spectra, averaged over vowels and modes of singing, also hold for each vowel and each mode of singing individually. To study this, we first performed the above-mentioned analysis for F_0 = 220 Hz for each vowel separately (by averaging spectra of each singer over nine modes of singing only). Generally, the same type of configuration of singer differences as presented in Figs. 5.1 and 5.2, was also found for each vowel individually. The extensiveness of the point configuration was, however, vowel dependent. This can be explained by the spectral variance, associated with spectral differences between singers for each vowel, which was greatest for /œ/ (330 dB^2) and /u/ (316 dB^2) and smallest for /ɑ/ and /a/ (both 179 dB^2).

In addition to the configurations, the interpretation of the spectral dimensions is of interest. The basis vectors of dimensions related to bass-tenor and alto-soprano differences are presented separately for each vowel and for all F_0 values in Appendix A. It showed that dimensions associated with singer differences had comparable spectral interpretations for the back vowels /u/, /ɔ/, /ɑ/, and /a/, as was also the case for the front vowels /i/, /y/, /e/, and /ɛ/. The vowel /œ/ gave intermediate results. Therefore, we present in Fig. 5.3, for F_0 =220 Hz (combined male and female data) the basis vectors of the first two dimensions and the grand-average spectrum for the representative vowels /a/ and /i/.

Dimension I typically describes vowel-dependent spectral variation. For the vowel /a/, a positive contribution of basis vector I indicates a relatively high sound level in the frequency bands with center frequencies of 0.4, 1.0, and 2.5 kHz, and a relatively low sound level in the frequency

bands with center frequency of 0.8, 1.6, and 4.0 kHz. Since a positive contribution of basis vector I is associated with bass singers (Fig. 5.1), this result can be interpreted as follows: Bass singers, relative to tenor singers, have (1) a lower first and second formant frequency, which has, however, a minor spectral effect, (2) a lower frequency of the high spectral peak (singer's formant), and (3) a lower cut-off frequency. The spectral effects above 1.2 kHz represent the major differences between male voice timbre types for the vowel /a/. The same type of description applies to the vowel /i/.

Dimension II is much less vowel dependent, except that the second basis vector for /i/ shows an additional weighting of the frequency of the second formant in the 1.6 and 2.0 kHz frequency bands. Furthermore, the interpretation of this basis vector follows the general description, also given for Figs. 5.1 and 5.2: a positive coordinate value, as in alto singers, indicates a slightly lower sound level of the fundamental (for the vowel /a/ only) and relatively high sound levels above 2.5 kHz.

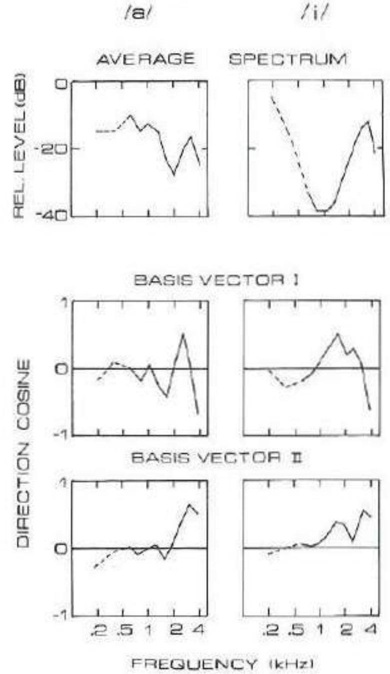

Fig. 5.3. Grand-average spectrum and profiles of basis vectors of the first two dimensions of a spectrum subspace, derived for the representation of spectral differences between singers, separately for the vowels /a/ and /i/ (F_0 = 220 Hz).

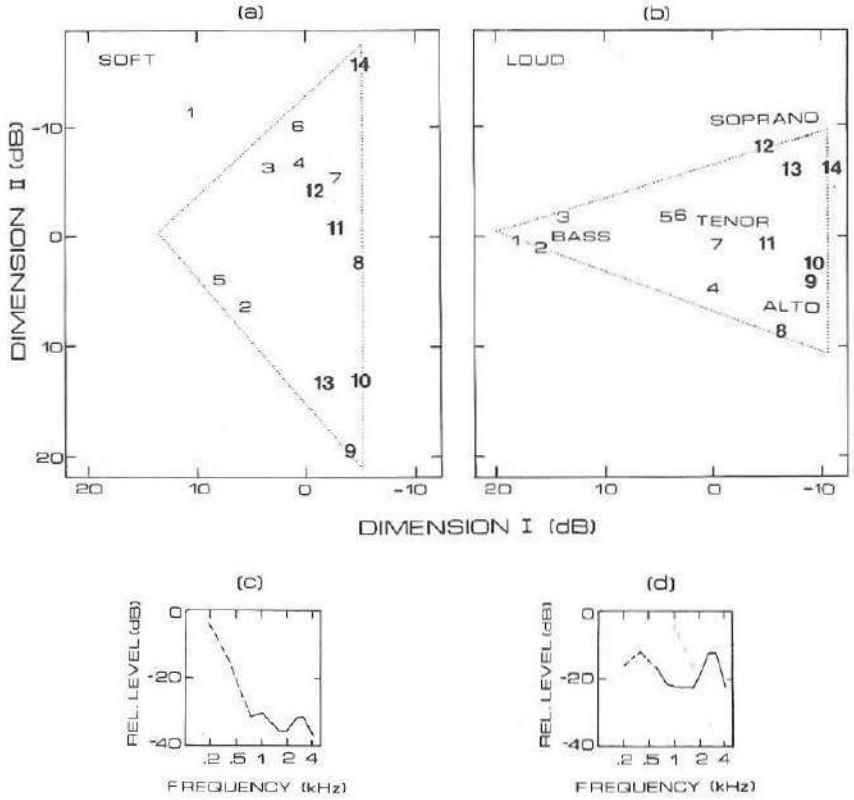

Fig. 5.4. Singer point configurations for the soft (panel a) and loud (panel b) modes of singing. Panels c and d present the respective grand-average spectra. Basis vectors of panels a and b are the same as presented in Fig. 5.1b. The numbering of the singers also follows Fig. 5.1c.

5.2.3 Singer differences for various modes of singing

To study the dependence of spectral differences between singers on mode of singing, we separately computed for each mode of singing (F_0 = 220 Hz) the spectrum of each singer averaged over nine vowels. This showed that for each mode of singing these average spectra could be described almost optimally with the same basis vectors as used for the general case (Fig. 5.1b). Furthermore, the configuration of points for the singers turned out to depend to a considerable extent on mode of singing. Interesting extremes were found for the soft and loud modes of singing. The corresponding singer configurations are shown in panels a and b of Fig. 5.4 in an I-II plane with the same basis vectors as in Fig. 5.1b. Panels c and d show the grand-average spectrum for both modes of singing. For the soft mode of singing, differences among both males and females are much larger in the second dimension than in the general case (Fig. 5.1a). For the loud mode of singing the opposite can be seen: differences between singers are concentrated more along the first dimension. This dimension shows clusters of (1) bass and baritone singers, with the exception of baritone singer 4, (2) (counter)-tenor singers, and (3) female singers. Variation in the second dimension is markedly reduced compared to the soft mode of singing, which affects especially the positions of the female singers. The other modes of singing showed intermediate results, whereas the same type of observations were obtained for F_0 = 392 Hz.

5.3 Discussion

5.3.1 Male voice classification

The main spectral differences among male voices, as well as average male-female differences, generally indicated lower formant frequencies for the lower voice classification (dimension I in Fig. 5.3). This result agrees qualitatively with observations by Cleveland (1977). However, in addition we found the greatest spectral differences between male voices in frequency bands with center frequencies higher than 1.2 kHz. This result agrees with the well-known observation for the speaking voice that the higher formants are most informative on personal (male) voice characteristics. For example, in the field of speaker recognition Sambur (1975) found F2 for front vowels and F3 for the back vowel /u/ among the most effective acoustic features for the identification of male speakers.

For professional male singers, spectral differences seem to come out especially in the frequency of the singer's formant, and can therefore be explained mainly in terms of lower vocal tract morphology (Sundberg, 1974): The singer's formant originates from an acoustical mismatch between the pharynx and the entrance of the larynx tube, which is promoted by a low larynx position and a wide pharynx, but not necessarily so. In this connection we remark that Cleveland 's (1977) suggestion that the same morphological differences between males and females may also distinguish between bass and tenor singers, is only correct for a comparison with average male-female differences, since our results showed that inter-individual spectral differences between female singers are of a different nature.

In addition, spectral differences between male singers were not exclusively represented in the first dimension; we also found some variation in the second dimension (tenor singer 6 in Fig. 5.1 and the phonations in the falsetto register in Fig. 5.2) of which we will discuss a possible glottal origin in the next section.

Furthermore, anticipating data presented in Chapter 7 and Appendix C, we state that there is a relation between male voice classification and the average overall SPL of vowels at the same F_0 values: On the average, bass singers sang about 3 dB louder than baritone singers, who, in their turn, sang about 3 dB louder than tenor singers. However, these systematic differences in overall SPL between singers do not seem to have important spectral effects: Spectral correlates of variation in overall SPL (Chapter 7) are of a different nature than spectral differences between male singers.

In relation to male voice classification, it is of interest to consider the results of the counter-tenor singer 7. Figs. 5.1 and 5.2 show that he approaches female singers very closely in dimension I. The counter-tenor, who originally had a baritone voice classification, probably sang with a raised larynx and in this way reduced his pharyngeal cavity. Possibly, the same holds for the falsetto phonation of singer 3, which had a lower coordinate value along dimension I than the phonations in the modal register. The spectra of phonations in falsetto register of singers 2 and 3 and of counter-tenor 7 differed mainly with respect to the relative levels of higher frequency bands (reflected in the coordinate values along dimension II). The average spectrum of the counter-tenor resembled those of modal tenor singing and had relatively high levels in the higher frequency bands. On the other hand, the counter-tenor sang on the average at a 6 dB lower overall SPL than the tenor singers. We may speculate that the counter-tenor in the falsetto register realizes a medial compression of the vocal folds comparable to modal register, with complete, and possibly even relatively long, glottal closure, resulting in a low amplitude of the glottal pulse, strong high harmonics in the output spectrum, and relatively low overall SPL. If this hypothesis is correct, it is of interest that the more medial compression is realized in the falsetto register, the more the apparent timbre type changes from soprano (singers 2 and 3) to mezzo-soprano

or alto timbre (counter-tenor 7). Source spectrum measurements will be needed to verify this hypothesis.

5.3.2 Female voice classification

One of our most remarkable observations was that female voices differed from each other in another way than male voices. Spectral differences between female voices were described primarily (1) by the sound level of the fundamental, which is the minor spectral effect, and (2) the sound level of the frequency bands with center frequencies of 3.16 and 4 kHz, which is the major spectral effect. For alto singers, the amplitude of the fundamental was lower and the sound level in the 3.16 and 4 kHz bands was much higher than for soprano singers. The latter observation is in conformity with the remark by Bartholemew (1934) that among female singers the singer's formant is found only in good altos.

An explanation of the spectral differences among female singers is difficult because 1/3-oct spectra are hard to interpret unambiguously in terms of sound production parameters. Still, we may recall the following two observations: (1) The main spectral differences among females are independent from those between males and from the average difference between males and females; therefore, the explanation given for differences between males, in terms of the effect of different vocal tract morphology, does not apply equally to females. This view is supported by the X-ray data from Dmitriev and Kiselev (1979), who found that the length of the vocal tract of female singers only discriminated for high sopranos. (2) The spectral difference between modal register and falsetto register in male singers has a strong component in the direction of spectral differences between females, whereas the register difference is generally accepted to be a laryngeal phenomenon (Colton, 1972; Hollien, 1972; Russo and Large, 1978; Sundberg and Gauffin, 1979). Both observations may indicate that glottal sound source characteristics rather than morphological differences underlie the main spectral differences between females. This view is supported by the observation by Karlsson (1984) that there is a wide range of glottal wave shape parameter values across female speakers - greater than any male-female differences.

Possibly, females vary in the extent to which the vocal folds are adducted, glottal closure duration, or the amplitude of the glottal pulse. These phenomena have their main effects on the sound level of the fundamental and on the sound levels of the high harmonics, as was shown theoretically by Fant (1981), and experimentally for males by Sundberg and Gauffin (1979) and Gauffin and Sundberg (1980). These effects can be seen during vocal intensity variation. However, no systematic relation was found between average SPL for female singers and voice classification (see Chapter 7 and Appendix C).

According to Rothenberg (1983), another possible explanation of the varying levels of high-frequency components may be found in the effect of varying pharyngeal morphology on the inertive part of the acoustic loading of the glottis. Especially the laryngeal ventricle, just above the glottis, has the greatest effect on the inertive loading of the glottis and a minimal effect on formant frequencies. However, Flach (1964) measured the lateral area of the laryngeal ventricle by means of X-rays and did not find a systematic relation with voice classification, whereas the variation among singers was large. Flach only mentioned that the size of the ventricle was larger in singers with a good voice than in subjects with voices of poor quality. The lack of a relation between voice classification and the size of the laryngeal ventricle equally falsifies a possibility proposed by van den Berg (1955), namely that the laryngeal ventricle (sinus Morgagni) behaves as a low-pass filter. If this were the case, our results would imply that alto singers have a small, and soprano singers a large ventricle. In itself this seems to be an unlikely option.

Some additional support for a glottal origin of the main differences between females can be found from the singer configurations for the soft and loud modes of singing (Fig. 5.4). These modes affected especially the variation in the second dimension, which displays the main female differences. The greater variation in dimension II for the soft mode of singing possibly has a plausible interpretation in the effect of greater variability among singers in the degree of glottal adduction. This interpretation may also explain the less systematic relation to voice classification of the coordinate value of the female singers in dimension II. In the opposite case of the loud mode of singing, the highly reduced variation in dimension II may be explained as the effect of about the same maximal glottal adduction for all singers. For both the soft and loud modes of singing there was considerable variation of maximally 20 dB in overall SPL among singers. There was, however, no systematic relation between overall SPL and coordinate values for both dimensions I and II.

After these arguments in favor of glottal differences, it cannot be

ignored that several problems remain unsolved. For instance, no formal theoretical evidence has been given yet which can explain the dramatic effects in the frequency bands of 2.5, 3.16 and 4 kHz (see the basis vectors of the individual vowels in Fig. 5.4). Fant (1979) presents some theoretical spectra resulting from varying the steepness of the glottal pulse at closure. The main effects are a constant sound level of the fundamental and an equal increase of all sound levels above about 1 kHz with increasing steepness of the glottal pulse. Of course, incomplete closure of the vocal folds when they are loosely adducted would also affect the higher cutoff frequency of a spectrum dramatically, and consequently the level variation in the highest frequency bands. Whereas such an effect is, for instance, often reported for the male falsetto register (see also Fig. 5.2), it is hard to believe that incomplete glottal closure occurs in professional female singing, let alone that incomplete glottal closure would be the basis of systematic differences between professional female singers.

Finally we have to consider the possibility of an articulatory explanation. For this, we may refer to the articulatory explanation Sundberg (1974) provided for the singer's formant. It is conceivable that the larynx position may be relatively low for alto singers, resulting in high sound levels above 3 kHz, and that the opposite is the case for soprano singers. However, this was not clearly confirmed by the X-ray measurements of Dmitriev and Kiselev (1979). Moreover, the suggested articulatory explanation also fails to explain the variability in the level of the fundamental between females. Still, articulatory differences between female singers may play a secondary part, which is demonstrated by the lowered second formant of front vowels for alto singers (Fig. 5.3).

From these considerations on the possible origins of the spectral differences between male and female singing we may hypothesize the existence of two main effects: morphological differences in the vocal tract, and glottal differences. Roughly speaking, for males and for the average male-female difference, the morphological differences prevail, for females the glottal differences do. If additional evidence for this view could be obtained from reliable glottal sound source measurements (up to 3-4 kHz), especially the puzzle of the female voice would be a step closer to its solution.

6 SPECTRAL CHARACTERISTICS OF DIFFERENCES BETWEEN MODES OF SINGING

The spectral characteristics of the modes of singing could be represented for each vowel in two dimensions. The modes of singing soft (pianissimo), light, neutral, free, straight, extra vibrato, and loud (fortissimo) differed mainly due to the spectral effect of vocal effort, and constituted a very dominant first spectral dimension. This dimension roughly weighted the slope of the spectrum. The second dimension mainly described the spectral differences between the dark and pressed modes of singing. A possible basis for the results in terms of glottal and morphological variation is discussed.

Part of a paper with R.Plomp, submitted to J.Acoust.Soc.Am.

6.1 Introduction

We investigated spectral properties of nine modes of singing: neutral, light, dark, pressed, free, soft, loud, straight, and extra vibrato. The last eight terms may be considered as pairs of adjectives with opposite meanings, referring to four, not necessarily independent, scales with "neutral" as their center: "light-dark", "pressed-free", "soft-loud", "straight-vibrato". Present knowledge of the acoustical aspects of these modes of singing is rather limited and fragmentary; this part of our study is devoted to a general spectral description over a wide F_0 range.

Little is known, for instance, of properties of light and dark phonations sung by the same singer. However, these terms are also used to characterize different timbres within the same voice classification. Therefore, if "light" and "dark" refer to general phonation types which can be chosen by a singer but which also distinguish between neutral phonations of different singers, we may make use of the following findings reported by Sundberg (1970, 1973). He measured formant frequencies and the source spectra of vowels sung by four bass singers with voice timbres varying between very dark and light. He found for the dark voice relative to the light voice (1) that four formants instead of five were needed to describe the vowel spectra, indicating a lower cutoff frequency, (2) a tendency towards lower formant frequencies, and (3) a less pronounced singer's formant due to smaller relative amplitudes of the higher source-spectrum partials. The first two findings are probably the acoustic effects of a lower larynx position and larger pharyngeal cavities for the dark voice, which were confirmed by X-ray measurements.

As a result of recent investigations into the vocal sound source, some data on pressed and free phonations are available. For pressed phonations, Sundberg and Gauffin (1979) found strong higher partials in the source spectrum relative to the fundamental. They suggested that this was associated with a high subglottal pressure and a high degree of laryngeal adduction, possibly as part of the total increase in muscle activity in the laryngeal region by which the larynx itself was also raised (Sundberg and Askenfelt, 1981). As a consequence of a raised larynx, formant frequencies may also increase in pressed phonations. Pressed phonation is, of course, not recommendable in singing, in contrast to free or "flow" phonation (Sundberg and Gauffin, 1979) which has just opposite characteristics.

The effect of vocal effort, the scale "soft-loud", on the spectra of spoken vowels is well-known and generally described by a rather stable sound level of the fundamental and a relative increase of the amplitudes of higher harmonics. Fant (1960, p.270) proposed as a rule-of-thumb for normal speech that, at a constant F_0, an increase of 10 dB in the level of the first formant is accompanied by an increase of only 4 dB of the level of the fundamental. Gauffin and Sundberg (1980) presented some data on the spectral effect of change in overall SPL for a singer and a non-singer. No substantial differences between the two subjects were found: as a modification of Fant's rule, the level of the fundamental was found to be stable up to medium vocal effort and to increase at the same rate as overall SPL for higher vocal effort. The levels of the formants increased much faster than did overall SPL for low to medium vocal effort and increased at about the same rate for higher vocal effort. This behavior was also observed with the level of the singer's formant by Sundberg (1973), Schultz-Coulon et al. (1979), and Hollien (1983). An explanation was proposed by Sundberg (1973) who found for two bass singers that, with increasing overall SPL, the amplitudes of source-spectrum partials above 1 kHz increased faster that those below 1 kHz.

Studies on the vibrato of sung vowels have always focused on the temporal aspects. No data seem to be available on spectra, averaged over at least one vibrato cycle, of vowels sung with various degrees of vibrato. However, it cannot be excluded that a certain rate and extent of vibrato, with the extremes of "straight" and "extra vibrato", does not only modulate fundamental frequency but is produced with typical properties of the average spectrum, too. We were able to look for these properties because our spectra were obtained as the averages from 30 10-ms spectra which included more than one normal vibrato cycle (which may vary between 120 and 200 ms in duration).

6.2 Results

6.2.1 Average spectra of modes of singing

In order to investigate the spectral variation due to different modes of singing, the average spectrum for each mode was computed. This average spectrum was derived from spectra of the realizations of a mode with the nine vowels by all singers. Since the spectral variance due to different

modes of singing did not vary much over F_0 (Chapter 3), we only present analyses for F_0 = 220 Hz for seven male singers and for F_0 = 392 Hz for seven female singers, which are about mid-range frequencies for both sexes.

A principal-components analysis of the nine mode spectra for these two sets revealed highly one-dimensional spectral variation. The first dimension of the computed spectrum space explained 95.2 % (males) and 96.1 % (females) of the total spectral variance due to modes of singing. Figs. 6.1 and 6.2 show the representations of the mode spectra in the I-II plane

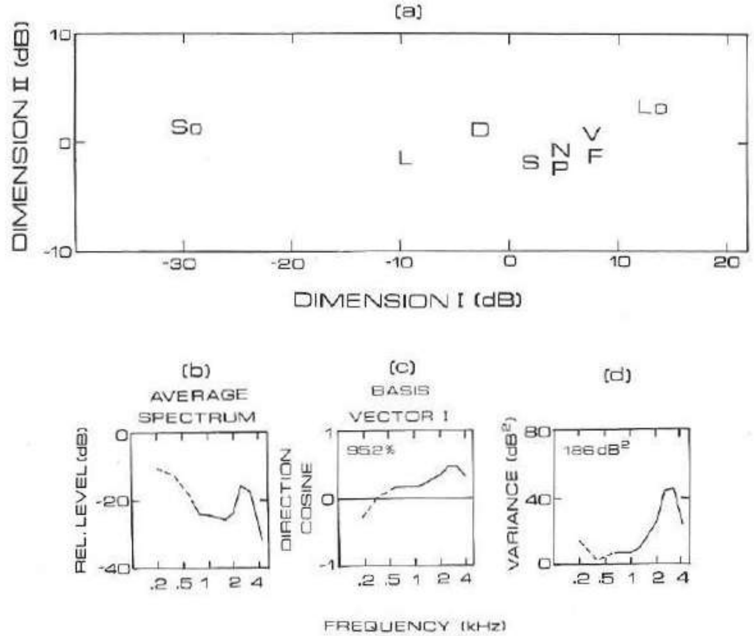

Fig. 6.1. Representation of spectral differences between modes of singing at F_o = 220 Hz for male singers.
(a) Configuration of modes of singing in a two-dimensional spectrum subspace (So=soft, L=light, D=dark, F=free, N=neutral, S=straight, P=pressed, V=extra vibrato, Lo=loud).
(b) Grand-average spectrum.
(c) Profile of the basis vector of the first spectral dimension.
(d) Distribution of spectral variance over frequency bands due to the factor "mode of singing".

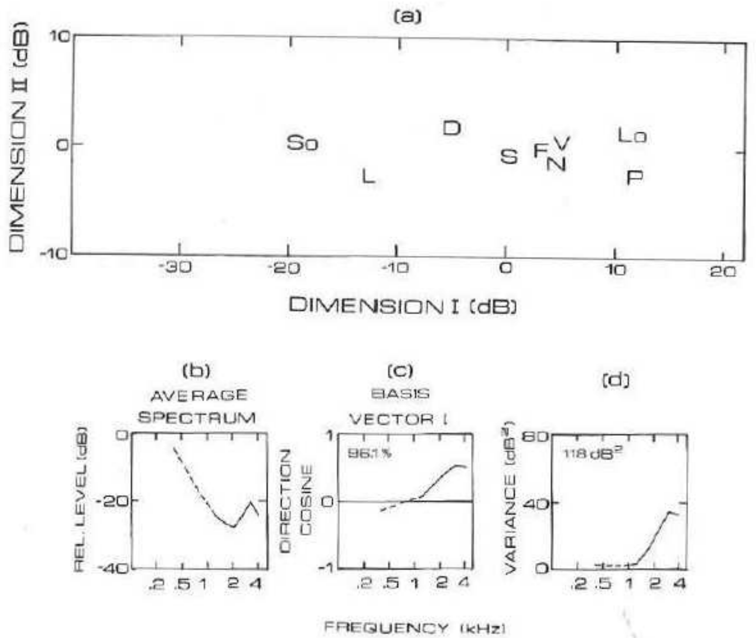

Fig. 6.2. Representation of spectral differences between modes of singing at F_o = 392 Hz for female singers. For an explanation of the panels see Fig. 6.1.

(panel a), the grand-average spectrum (panel b),the profile of the basis vector of the first dimension (panel c), and the contributions of each frequency band to total spectral variance (panel d). It turns out that the soft and loud modes of singing, representing minimum and maximum vocal effort, are extremes of the single important spectral-mode dimension. For male singers the profile of the first basis vector (Fig. 6.1) indicates that from soft to loud singing, passing all other modes of singing, the sound level of the fundamental decreases relatively, while the levels of the frequency bands between 2 and 4 kHz increase relatively. For female singers spectral variation is limited to a relative increase of the levels of the frequency bands above 1.6 kHz.

The modes of singing neutral, free, straight, and extra vibrato gave almost equal results, which indicates that (1) neutral and free are different descriptions of the same, comfortable way of singing, and (2) that absence or enhanced presence of vibrato does not influence the average

spectral composition. As a consequence, there remain six spectrally different singing modes, in this order: soft, light, dark, neutral, pressed, and loud. For a further exploration of their positions we have to take into account the average overall SPL of each mode. For the male singers these values were 77, 84, 91, and 100 dB SPL for the singing modes soft, light, neutral, and loud, respectively; for females the corresponding values were 82, 87, 92, and 99 dB SPL. This order of SPL values is the same as that of the coordinate values along dimension I. The dark and pressed modes of singing show a different behavior in that their average overall SPL value is about equal to the value of the neutral mode for both males and females, whereas with both male and female singers the position of the dark mode along dimension I is lower than that of the neutral mode, and the position of the pressed mode of female singers is higher. The consequences of these results will be discussed in Section 6.3.

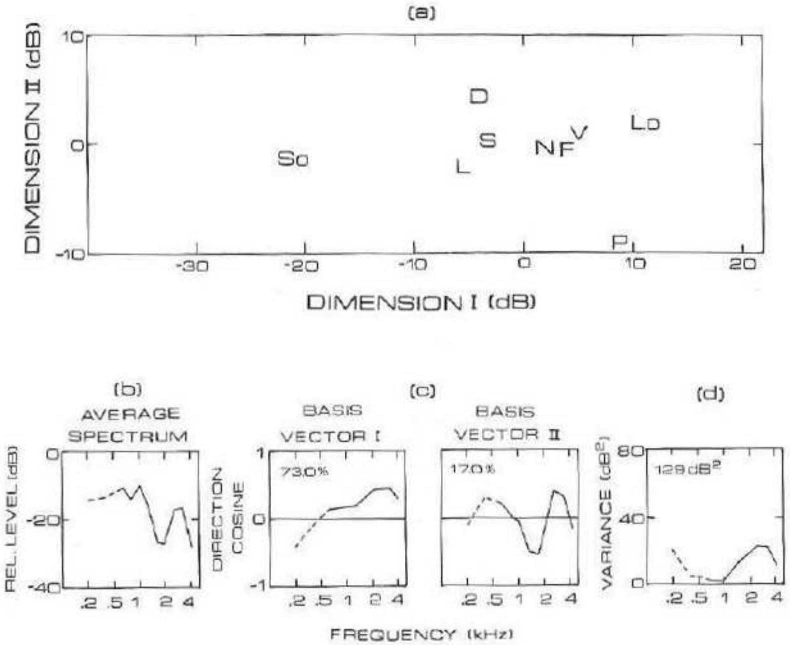

Fig. 6.3. Representation of spectral differences between modes of singing for the vowel /a/. Results for male singers at F_o = 220 Hz. For an explanation of the panels see Fig. 6.1. Panel c also presents the profile of the basis vector of the second dimension.

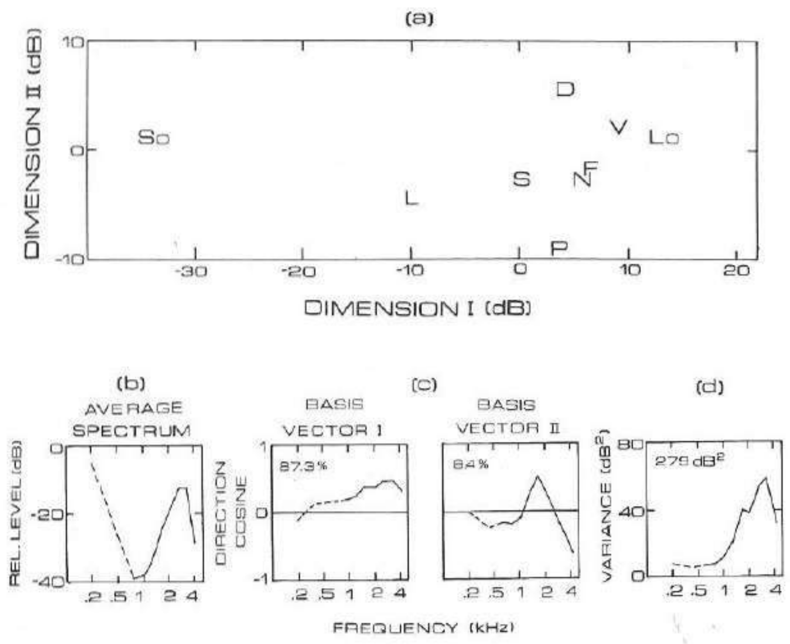

Fig. 6.4. The same presentation as in Fig. 6.3 for the vowel /i/.

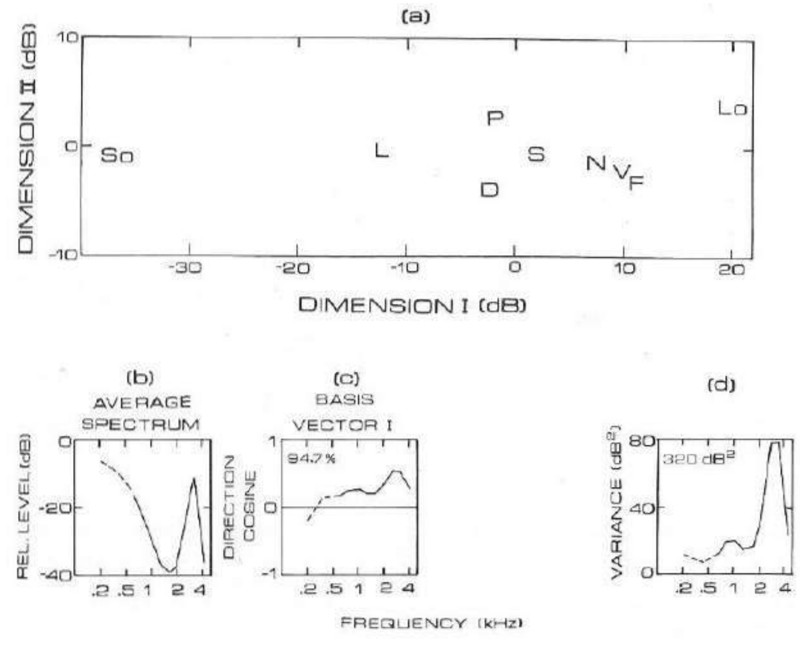

Fig. 6.5. The same presentation as in Fig. 6.3 for the vowel /u/.

6.2.2 Differences between modes of singing for various vowels

Up to now we have considered the one-dimensional spectral main effect of the modes of singing. For a further differentiation we have to study vowel-dependent spectral differences. Figs. 6.3, 6.4, and 6.5 show in panels a, for the male vowels /a/, /i/, and /u/, the representations of the nine modes of singing in the first two dimensions of the spectrum space, derived by principal-components analysis for each vowel individually. In addition, these figures present the standard information for their spectral interpretation in panels b, c, and d. Furthermore, in Fig. 6.6 the average spectra of the soft, loud, pressed, and dark modes of singing are plotted for the three vowels. The following differences between vowels can be observed:

(a) The variation along dimension I increases considerably from /a/, via /i/ to /u/. This effect is reflected in the increase in spectral variance due to modes of singing which increases from 129 dB^2 for /a/ and 279 dB^2 for /i/, to 320 dB^2 for /u/.

(b) The first dimension (see the basis vector in panels c) is comparable for all three vowels. Its spectral effect can also be seen in Fig. 6.6 as the difference between the average spectra of the soft and loud modes of singing.

(c) The second dimensions for /a/ and /i/ especially differentiate between the dark and pressed modes of singing. For /u/ the results are almost one-dimensional (94.7 % of the variance); for this vowel the small variation in the higher dimensions did not allow a further explanation.

(d) The spectral meaning of the second dimension is typically vowel dependent. For /a/, the profile of the basis vector (panel c) shows that a positive coordinate value (dark voice) implies (1) higher sound levels of the lower frequency bands of 0.4 and 0.63 kHz, which indicates lower F_1 and F_2, and (2) higher sound levels in the frequency bands of 2.5 and 3.16 kHz, indicating a more pronounced spectral peak (singer's formant). For /i/ the profile of the second basis vector shows for a dark voice a shift of the high spectral peak to lower frequencies, rather than an increasing level of this peak. The spectral meaning of the second dimension can also be seen in Fig. 6.6 as the major part of the difference in the average spectra of the pressed and dark modes of singing for /a/ and /i/.

The results presented in this section for a single fundamental frequency for male and female singers generally hold for all F_0 values. Complete data can be found in Appendix A.

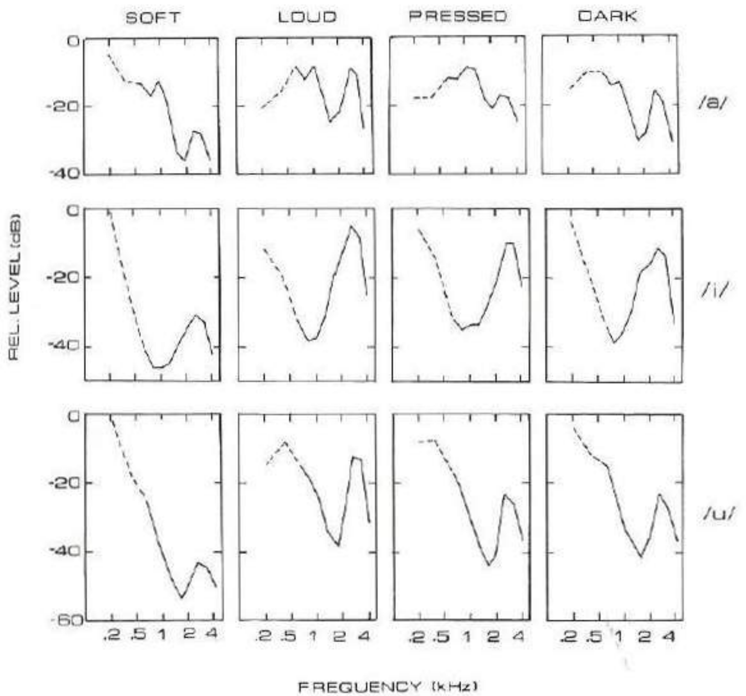

Fig. 6.6. Average spectra for male singers with F_0 = 220 Hz for the vowels /a/, /i/, and /u/, and the modes of singing soft, loud, pressed, and dark.

6.3 Discussion

6.3.1 The dimension soft-loud

We found that systematic differences between modes of singing of a vowel could be represented quite well in two dimensions of the spectrum space for all F_0 values. The most important first dimension referred to the spectral effect of vocal effort and could describe the main spectral effects of the singing modes soft, light, neutral, free, straight, vibrato, and loud. If we interpret the profile of the basis vector of the first dimension (Figs. 6.1b-6.4b) for the moment as being entirely the consequence of variation in the glottal sound source, this profile is, up to 1.5 kHz, in agreement with theoretical data of Fant (1981) and experimental data of Sundberg (1973). With respect to increasing overall SPL we may expect: (1) relatively little variation in the sound level of the fundamental, and

(2) about the same increase of the sound level of harmonics beyond F_1. Translated in terms of 1/3-oct spectra normalized for overall SPL (which is, generally speaking, mainly determined by the sound level of F_1), this implies (1) a relative reduction in the sound level of the fundamental, especially when F_0 and F_1 are distant, and (2) an increase in sound level for frequency bands with center frequencies beyond F_1, which is equally large for all frequency bands. These acoustic effects are the results of a steeper glottal pulse, due to increased adduction of the vocal folds (see also Gauffin and Sundberg, 1980). No supporting source-spectrum data are available for the extra level increase in the frequency region of 2-3.16 kHz with increasing overall SPL. It is conceivable that during low vocal effort there is no complete closure of the vocal folds; this would have a great influence on the cut-off frequency of the spectrum, resulting in large level variations in the higher frequency bands. It is also this frequency region in which the vowel-dependent spectral effect of an increase in vocal effort comes out most clearly (see also Chapter 7).

From the glottal point of view the deviating positions of pressed and dark phonations along dimension I for female singing (Fig. 6.2a) and for the vowel /a/ for male singing (Fig. 6.3a) can be associated with source-spectrum characteristics of "pressed" and "flow" phonations, measured by Sundberg and Gauffin (1979). Firstly, they found that the amplitude of the glottal pulse is relatively small for pressed phonations and relatively large for flow phonations. Therefore, other things being kept constant, overall SPL will also be relatively low for pressed phonations and relatively high for flow phonations. When, however, the same overall SPL is obtained for the neutral, pressed, and dark modes of singing, this can only be due to a second cause, namely the degree of adduction of the vocal folds. With a relatively high degree of adduction for the pressed phonation and a relatively low degree of adduction for the dark phonation the same overall SPL can be obtained. If we assume that especially the spectral effect of vocal-fold adduction is present in dimension I, this explains the deviating positions of pressed and dark phonation, provided that dark singing is produced as flow phonation. Reversing the argument, our data (Figs. 6.1, 6.4, 6.5) then suggest that the difference between pressed and dark or flow phonation was less outspoken in male singing, and for the vowels /i/ and /u/.

After these considerations in terms of the glottal sound source we now consider the possibility of an articulatory interpretation of the results. In Appendix B it is shown that for phonations of a bass singer (F_0 = 98 Hz), the variation described by dimension I can be interpreted as the effect of a specific variation in the first two formant frequencies. This articulatory explanation is, however, not satisfactory because of the unrealistically large formant frequency variation required. Nevertheless, it is probable that articulatory variation associated with vocal effort, for instance the degree of mouth opening, plays a part in the results.

6.3.2 The dimension pressed-dark

The basis vector II in Figs. 6.3 and 6.4 generally describes lower formant frequencies for the dark mode of singing and higher formant frequencies for the pressed mode of singing. This dimension may be interpreted, therefore, as representing the spectral effect of pharyngeal volume, influenced by the height of the larynx, which can also be seen in tomograms given by Sundberg (1973). Since the dark and pressed modes of singing also vary with respect to dimension I, we may combine the results and describe the dark mode of singing as flow phonation with a wide pharyngeal cavity, and the pressed mode of singing as pressed phonation combined with a high larynx position, the latter in accordance with Sundberg and Askenfelt (1983). In this connection it should bementioned that pressed vowels were sometimes realized in two different ways: most singers noted that, apart from an overall increase in muscular tension, they raised their larynx and produced a timbre type also known as "Knödel". Some male singers, however, lowered their larynx and produced an extremely widening of the pharyngeal cavity for the timbre type "poitrinated" which turned out to be closely related to dark sung vowels (not separately shown in the presented figures).

6.3.3 Some perceptual implications

From the perceptual point of view, our results indicate that variation in vowel timbre due to mode of singing may be perceived as a two-dimensional phenomenon for each vowel. Studies by Von Bismarck (1974a,b) showed "sharpness" to be the most important attribute of the timbre of harmonic complex tones. He found that sharpness was related to the relative importance of higher harmonics, which corresponds well with the spectral interpretation of dimension I. In this respect it is of interest that the profile of basis-vector I is not very vowel dependent, allowing a

general interpretation of sharpness. On the other hand, the spectral differences between dark and pressed phonations were vowel dependent and therefore probably not traced as a prominent factor in Von Bismarck's experiments.

6.3.4 Relations between the factors "modes of singing" and "singers"

The spectral effects of the factors "singers" and "modes of singing" have much in common. Firstly, in Chapter 3 it has been shown that the dependence of spectral variance on fundamental frequency is similar for both factors: spectral variance is greatest in the higher frequency bands which are not much affected by increasing F_0. Secondly, and more specifically, the basis vector of the first two spectral dimensions of both factors are quite comparable: The first mode dimension (soft-loud, Figs. 6.1b and 6.2b) resembles the second singer dimension quite well (soprano-alto, Figs. 5.1b and 5.2b), and the second mode dimension (pressed-dark, Figs. 6.3b and 6.4b) resembles the first singer dimension (females-males, tenor-bass, Fig. 5.3). Although this comparison is not optimal in all cases, at least it expresses a strong tendency. We may speculate that this tendency may be extended to the production level: one dimension which mainly expresses the spectral effect of morphological differences (male vs. female, tenor vs. bass voice timbre type) and variation (pressed and light vs. dark mode of singing), and a second dimension which mainly expresses the spectral effect of glottal differences (soprano vs. alto voice timbre type, falsetto vs. modal register) and variation (soft vs. loud mode of singing). Comparing the size of the spectral variation in all dimensions, it can be seen in Figs. 5.1a, 5.2a, 6.3a, 6.4a, and 6.5a that the spectral difference a singer makes between a soft and a loud phonation is larger than the maximal spectral difference between female singers. The spectral difference between pressed and dark is somewhat smaller than the maximal spectral difference among male singers or between male and female singers. This comparability of spectral variations demonstrates that a singer has many possibilities to adapt his voice timbre to the timbre of other voice types, or to the esthetic requirements of artistic performance.

7 THE OVERALL SOUND-PRESSURE LEVEL AND ITS SPECTRAL CORRELATES

For all F_0 values investigated, the variation in overall SPL of sung vowels was largely due to the effect of different modes of singing, especially of singing soft vs. loud. Variation due to differences between singers came second. The spectral correlates of variation in overall SPL were spectral-slope like. An explanation is proposed for the result that the spectral effect associated with overall SPL was vowel dependent, i.e. much larger for /u/ than for /a/.

7.1 Introduction

For the same fundamental frequency, vowels can be sung with a dynamic range of more than 40 dB by an individual singer. In Section 7.2 we present the variation in overall SPL due to differences between singers, vowels, and modes of singing. Furthermore, variation in overall SPL of sung and spoken vowels is not simply a gain factor but has specific spectral effects. This has already been demonstrated by the spectral differences between the soft and loud modes of singing. In Section 7.3 we will explore the relation between vowel spectrum and overall SPL in a formal way for all F_0 values investigated.

7.2 Variation of overall SPL due to differences between vowels, singers, and modes of singing

The recording procedure of the sung vowels included a calibration procedure which allowed the determination of the overall SPL of each vowel at a microphone distance of 0.3 m. On these data, separately for each F_0 value, analyses of variance were performed to investigate the extent to which SPL varied due to the main effects and interactions of the factors "vowels," "singers," and "modes of singing".

Table 7.1 presents the variance due to the various factors and interactions. Most factors and interactions were significant beyond the 0.01 level. It can be seen that for most F_0 values the main effect of the factor "modes" explains more than half of the total variance and is therefore the most important source of variation in SPL. The second important factor is the singer-related difference in SPL. For males, this effect proved to be systematically related to voice classification. On the average, bass singers sang about 3 dB louder than baritone singers, who, in their turn, sang about 3 dB louder than tenor singers. Complete data of average values of overall SPL for singers, modes of singing and vowels are presented in Appendix C. For females no such relation was present. The relatively large variance was due to mezzo-soprano 11, whose SPL was, for all F_0 values, 8 dB lower than the average. For the same F_0, males sang on the average 6.4 dB (F_0 = 220 Hz) and 9.3 dB (F_0 = 392 Hz) louder than female singers. In modal register males sang on the average 9.3 dB louder than in the falsetto register. It is noteworthy that overall SPL was more constant in the falsetto register than in the modal register. Furthermore, there was a

Table 7.1. Variance (dB^2) in overall SPL due to main effects and interactions of the factors "modes of singing" (M), "singers" (S), and "vowels" (V), as a function of F_o. All main effects and interactions are significant beyond the 0.01 level, except those marked with an asterisk. For computational details see Table 3.2.

F_o (Hz)	Males 98	Males 131	Males 220	Males 392 modal	Males 392 falsetto	Females 220	Females 392	Females 659	Females 880
Modes	29.7	37.8	35.9	34.3	32.7	29.2	19.4	16.0	56.7
Singers	5.9	8.7	10.6	11.2	1.3	16.0	16.1	13.3	2.2
SxM	7.2	6.7	7.3	9.1*	5.8	6.2	6.1	6.4	5.0*
Vowels	2.1	1.7	1.7	0.4	3.1*	1.7	0.9	0.4	0.1*
SxV	0.6*	0.7	1.1	1.5	0.4	0.8	1.2	0.8*	0.3*
MxV	0.8	1.0	1.1	1.5	3.2	0.9	0.5	0.3	1.3
SxMxV	1.9	2.8	2.5	2.6	3.0	1.8	1.8	1.6	1.3
Total	48.2	59.4	60.2	60.6	49.5	56.6	46.0	38.8	66.9

relatively large interaction between singers and modes of singing, i.e. the dependence of overall SPL on modes of singing differed from singer to singer.

The main effect and all interactions of the factor "vowels" were much smaller and several were not significant at the 0.01 level. Nevertheless, a systematic tendency could be observed with respect to average SPL for different vowels (see Appendix C). Up to F_o = 220 Hz, overall SPL of the vowels /a/ and /ɑ/ was about 3 dB higher than that of the vowels /i/, /u/ and /y/, and the other vowels had SPL values in between; for F_o = 392 Hz, in male falsetto register and in female singers, this relation was just reversed: /i/, /u/, and /y/ had the higher overall SPL. For still higher F_o values no systematic differences between vowels were found.

The factor "modes of singing" was investigated in some more detail. Figure 7.1 shows SPL averaged over singers and vowels as a function of F_o, with mode of singing as the parameter. Since differences in SPL between the modes neutral, dark, free, pressed, straight, and vibrato appeared to be small, their values are presented as a range indicated as neutral. In this way four distinct stages could be distinguished: soft, light, neutral, and loud, at distances of, on the average, 6.6 dB, which means roughly a doubling in loudness. There was a monotonous increase of SPL with rising F_o, with the exception of the vowels sung by males in falsetto register and for the softly sung vowels at F_o = 880 Hz.

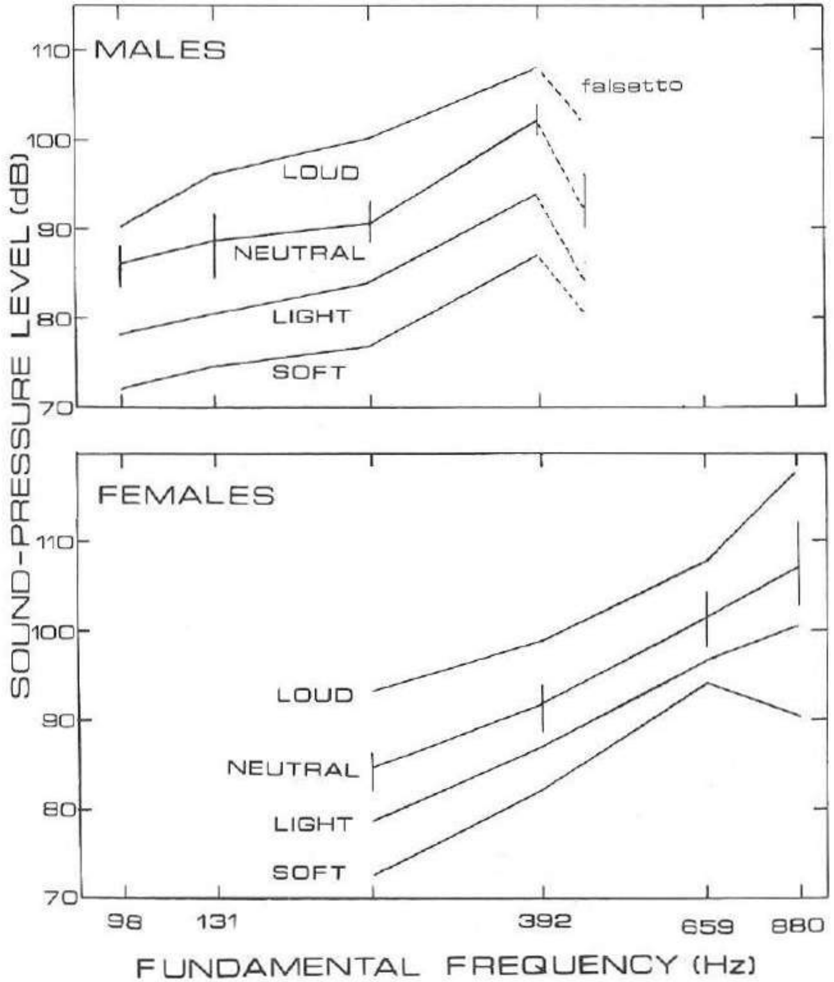

Fig. 7.1. Overall SPL for modes of singing averaged over vowels and singers as a function of F_o, for male and female singers separately. The line labeled "neutral" combines the results of the modes of singing neutral, dark, free, pressed, straight, and extra vibrato; the range of the average data for these modes is represented by vertical dashes.

7.3 Spectral correlates of overall SPL

The procedure that was used to find spectral correlates of overall SPL deviates from the general analyses given in Chapters 4 to 6 for vowel spectra. Firstly, we used the original, unnormalized, spectra; this implies the use of absolute sound levels in frequency bands as a function of overall SPL. Secondly, we separately investigated for each frequency band the relation between the sound level in that band and overall SPL by means of linear regression analysis. If the regression coefficient equals one, the sound level in the corresponding frequency band increases equally with overall SPL; if the regression coefficient is smaller, the sound level increases more slowly, and if the regression coefficient is greater the sound level increases faster than does overall SPL. At each F_0 value investigated, these analyses were performed separately for each vowel. Therefore, each analysis involved the spectra of the nine modes of singing, sung by all participating singers. For example, Fig. 7.2a shows the SPL of the vowel /a/ (F_0=392 Hz) in the frequency band of 3.16 kHz as a function of overall SPL. Data from spectra of nine modes of singing, realized by seven female singers, are used.

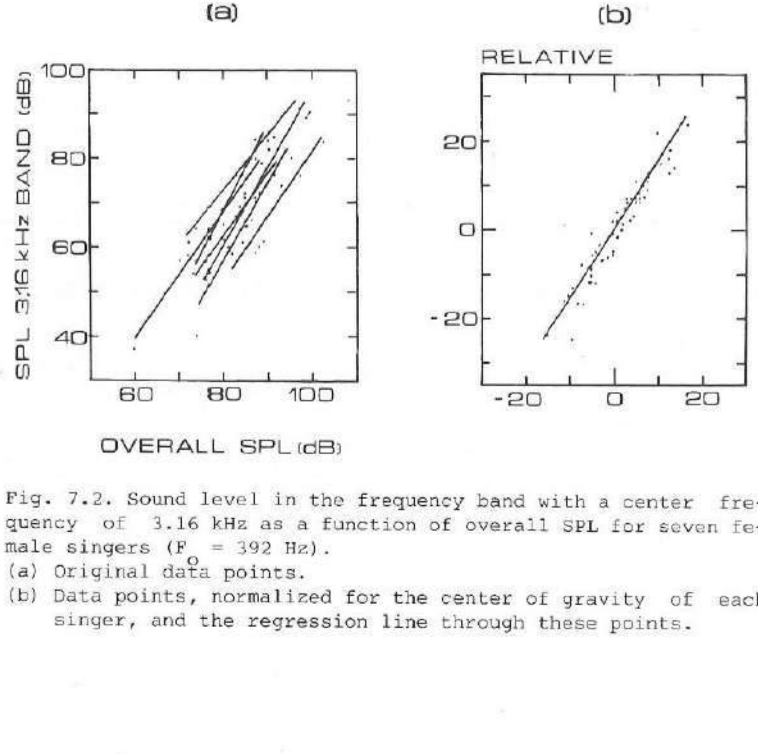

Fig. 7.2. Sound level in the frequency band with a center frequency of 3.16 kHz as a function of overall SPL for seven female singers (F_0 = 392 Hz).
(a) Original data points.
(b) Data points, normalized for the center of gravity of each singer, and the regression line through these points.

To avoid the influence of systematic level differences between singers on the regression line through all data points, we first normalized the data points of each singer for their combined center of gravity. Fig. 7.2b shows the normalized data and the regression line (regression coefficient = 1.49; correlation coefficient = 0.94). Using normalized data, correlation coefficients exceeded 0.90 in most cases. However, for the frequency band of the fundamental the correlation coefficient was lower, about 0.70, for F_0 values up to 392 Hz (males). Regression coefficients are shown graphically in Fig. 7.3; they are only given for the vowels /a/ and /u/ since these vowels represented the extreme cases. The regression coefficients for successive frequency bands are connected. The following observations can be made.

(1) If F_0 is lower than about 392 Hz, the sound level increases equally with overall SPL only in the frequency band in which the first formant is located (center frequency of 0.307 kHz for /u/, 0.5 - 0.63 kHz for /a/). This is not surprising since the level of the first formant determines, in most cases, overall SPL. For females (except F_0 = 220 Hz), the sound level of the first two or three harmonics increases equally with overall SPL. For these F_0 values, the first harmonic often determines overall SPL (see, for instance, the average spectra for these subsets in Fig. 4.7)

(2) With F_0 values up to about 392 Hz, the sound level in the frequency bands lower than the first formant frequency increases more slowly than overall SPL. In most cases the increase in the sound level of the frequency band of the fundamental is lowest and equal for all vowels, although the variability is somewhat larger than in other frequency bands. For F_0 values in the range of normal speech, the regression coefficient for this band has a value of about 0.4, which is in conformity with the rule given by Fant (1960): a 4 dB level increase of the fundamental when overall SPL increases by 10 dB. The increase in the frequency bands between the fundamental and F_1 is vowel dependent, probably because of differences in F_1.

(3) Still more vowel dependent is the regression coefficient for frequency bands higher than F_1. For frequency bands with a center frequency higher than the one in which the first formant is located (up to about 2 kHz), the value is roughly constant for each separate vowel, but at a different level for different vowels. The extra increase for /u/ is about double or triple that of /a/.

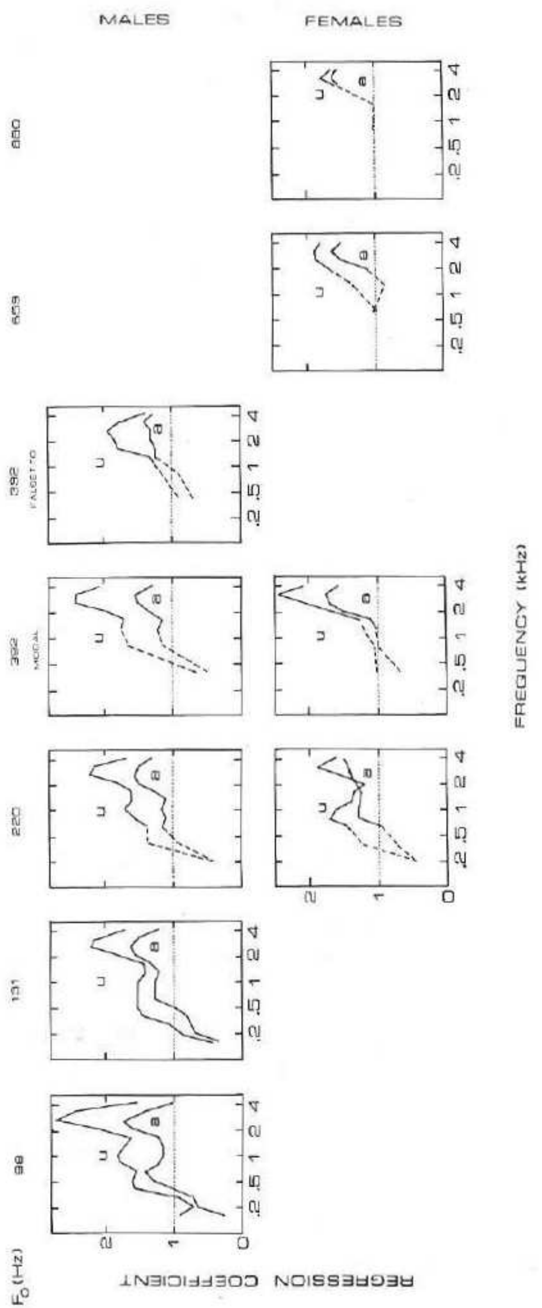

Fig. 7.3. Regression coefficient between the level in a frequency band and overall SPL, for the vowels /a/ and /u/. For all F_0 values investigated, the upper panels present the results for the male singers, the lower panels those for the female singers.

(4) This vowel-dependent relation between SPL in a frequency band and overall SPL persisted for the frequency bands between 2 and 4 kHz. For these frequency bands the extra increase in sound level is very large. At F_0 = 98 Hz the sound level of /u/ in the frequency band with a center frequency of 2.5 kHz increases by 27 dB with a 10 dB increase of overall SPL, whereas for /a/ this is 15 dB. The corresponding correlation coefficients have typical values of 0.91 and 0.97. For the vowel /a/ the value of the regression coefficients for these frequency bands is approximately the same over all F_0 values. The values for the vowel /u/ decrease with increasing F_0 to the same value as for /a/.

7.4 Discussion

The results presented above give an explicit relation between overall SPL and its spectral correlates. Of course, there is a close relationship between these results and the spectral correlates of modes of singing described previously, since it has been shown that modes of singing have the greatest contribution to variation in overall SPL. This close relationship is exemplified by the correspondence between the profiles of the basis vectors of dimension I of the modes of singing (Figs. 6.3a and 6.5a) and the profiles of the regression coefficients presented in Fig. 7.3. It should be noted that the series of regression coefficients for subsequent frequency bands can be considered as a vector. Transformation to a basis vector includes: (1) substracting the value one, which is the consequence of the use of normalized spectra in previous spectral descriptions, and (2) scaling with a constant to obtain a unit-length vector.

We are inclined to believe that the main origin of the spectral effects of variation in overall SPL is the degree of adduction of the vocal folds, necessary to control varying subglottal pressure. Because of the correspondence to the description of modes of singing we refer for this discussion to Section 6.3.

Here we want to discuss the interesting vowel-dependent character of the spectral correlates of overall SPL (Fig. 7.3). The first thing we note is that the increase in the level of the fundamental is not vowel dependent. For low F_0 values the level increase is smaller than the increase in overall SPL; for high F_0 values, when F_0 comes in the neighborhood of F_1 (thus earlier for /u/ than for /a/), the increase of the sound level of the fundamental entirely determines the increase in overall SPL. Sundberg and

Gauffin (1979) showed a high correlation between the sound level of the fundamental and the amplitude of the glottal pulse. Therefore, we may conclude that the amplitude of the glottal pulse increases with overall SPL independently of the phonated vowel.

An increase of the amplitude of the glottal pulse is automatically accompanied by an increase in steepness of the glottal pulse if glottal closing time (between maximum amplitude and closure) is constant. Acoustically, this increase has a tilting character: the sound levels of the higher harmonics are increased most. Therefore this effect may explain the general shape of the spectral correlates of SPL in Fig. 7.3.

For the lower F_0 values, the level of the first formant determines overall SPL. When the amplitude of the glottal pulse is raised equally for the vowels /u/ and /a/, the effect on the level of the harmonics underlying F_1 will be different because the accompaning tilt of the spectrum will affect the harmonics underlying F_1 of /a/ more than those of the lower F_1 of /u/. Only a steeper glottal pulse for /u/ can provide a sufficiently high level of the harmonics underlying F_1 and thus provide an equally high overall SPL as for /a/. This steeper glottal pulse can be obtained, for instance, by higher medial compression, resulting in more adduction of the vocal folds and a shorter closing time of the glottis. If this were the case there would be a correlation between vocal fold adduction and the first formant frequency for the realization of a certain value of SPL.

Regardless of its precise origin, we observed far more spectral-slope-like variation for /u/ than for /a/. We may bring this in relation with a perception experiment by Carlson and Granström (1976) on the detectability of changes in spectral slope of vowels. They showed the threshold for /u/ was much higher than for /a/ and /i/, which indicates an interesting balance between production and perception of vowels with respect to spectral slope.

8 THE SOUND LEVEL OF THE SINGER'S FORMANT

The relative sound level of the "singer's formant", measured in the 1/3-oct band with a center frequency of 2.5 kHz for males and of 3.16 kHz for females, has been investigated with respect to variation due to singers, modes of singing, vowels, overall sound-pressure level, and fundamental frequency. Variation in the sound level of the singer's formant due to differences among male singers was only small (4 dB), the factors vowels (16 dB) and fundamental frequency (9-14 dB) had intermediate effect, while the largest variation was found for differences among female singers (24 dB), between modes of singing (vocal effort) (23 dB), and in overall sound-pressure level (more than 30 dB). In spite of this great potential variability, for each mode of singing the sound level of the singer's formant was remarkably constant up to F_0 = 392 Hz, due to adaptation of vocal effort. This may be explained as the result of the perceptual demand of a constant voice quality. The definition of the singer's formant is discussed.

8.1 Introduction

Bartholomew (1934) was the first to describe, as a physical attribute of a "good" voice quality in professional male singing, a pronounced high frequency peak in vowel spectra, the "2800", which was said to add "ring" to the voice. Since then, this spectral peak has been investigated by several researchers (Winckel, 1953; Rzevkin, 1956; Sundberg, 1974; Schultz-Coulon et al., 1979; Hollien, 1983) and is presently known as the "singer's formant". Sundberg (1974) showed that the singer's formant consists of a clustering of the third, fourth, and fifth formants, which is the effect of an acoustical mismatch between the pharynx and the entrance of the larynx tube. Such a mismatch may be promoted by a low larynx position and a wide pharynx. Although Bartholemew (1934) already recognized that the sound level of the singer's formant varied with vocal intensity, fundamental frequency, and voice classification, up to now no systematic study has been undertaken to establish the contribution of all these factors in one experimental design. This chapter aims to present these contributions.

8.2 Method

The sound level of the singer's formant was measured in the frequency band with a center frequency of 2.5 kHz for males and of 3.1 kHz for females. In the great majority of conditions, except a few at very high F_0 values, the sound levels in these frequency bands were the highest ones above 1.6 kHz. For further analysis the difference value from overall SPL, L_{sf}, was used. According to these measurement conditions any phonation, sung or spoken, has a singer's formant by definition, which is contradictory to what the term was intended to mean and therefore debatable. Nevertheless, we do not want to discuss the definition of the singer's formant at this point; after the presentation of our data such a discussion will be presented in Section 8.4.

Although it may seem attractive to follow the same approach for the analysis of L_{sf} as in Chapter 7 for overall SPL, an analysis of variance could not be applied to L_{sf}. The reason for this is that overall SPL was a dependent rather than an independent variable, because recordings were not made at fixed values of overall SPL. It turned out that overall SPL and L_{sf} were strongly correlated; therefore we applied regression analyses to L_{sf}

as a function of overall SPL, as demonstrated in Fig. 8.1 for the set of 567 vowels sung at F_0 = 220 Hz by seven female singers. Fig. 8.1a shows each singer's regression line through all her L_{sf} data points (9 vowels x 9 modes of singing). It can be seen that in this subset (1) L_{sf} is dependent on voice classification: for the same overall SPL alto singers (8,9,10) have a much stronger singer's formant than soprano singers (13,14), (2) the relative relationship between L_{sf} and overall SPL does not vary much among singers since the regression lines are approximately parallel, and (3) ranges of overall SPL vary among singers.

The singer-specific L_{sf} was defined as the value of the regression line at the average overall SPL (in this case 83.4 dB SPL). This definition is somewhat arbitrary because of some variation in the slope of the regression lines but it seems to be the best choice. Subsequently the L_{sf} data were singer-normalized by correcting them for the corresponding singer-specific L_{sf}.

With these singer-normalized L_{sf} data we subsequently studied the effect of vowels on L_{sf} (Fig. 8.1b). By analogy with the description of singer differences, regression lines are drawn for each vowel through all the singer-normalized L_{sf} data points as a function of overall SPL (9 normalized singers x 9 modes). We notice (1) comparable ranges of overall SPL for all vowels, (2) the steepest slope for the vowel /y/; the shallowest slopes for the vowels /a/ and /ɑ/, and (3) the lowest L_{sf} for the vowel /u/. Despite the differences in slope we again defined the vowel-specific L_{sf} as the value of the regression line for a vowel at average overall SPL (83.4 dB). Subsequently the L_{sf} data were vowel-normalized by correcting them for the corresponding vowel-specific L_{sf} (Fig. 8.1c).

In Fig. 8.1c the regression line through all 567 singer-normalized and vowel-normalized L_{sf} data points is drawn as a function of overall SPL, and the average positions of the different modes of singing are indicated. In view of their close distance to the regression line it may be concluded that, on the average, there is no specific relation between modes of singing and L_{sf} other than that governed by overall SPL. To show the variability in L_{sf} for a mode of singing, the contour of the data points for the soft and loud modes is given as a dotted line in Fig. 8.1c. Both modes cover substantial ranges of overall SPL, and they even overlap. This shows that variation of L_{sf} within a mode of singing can be explained in part as the effect of variation in overall SPL. Besides this, for a single value of overall SPL a standard deviation in L_{sf} of about 5.5 dB remained.

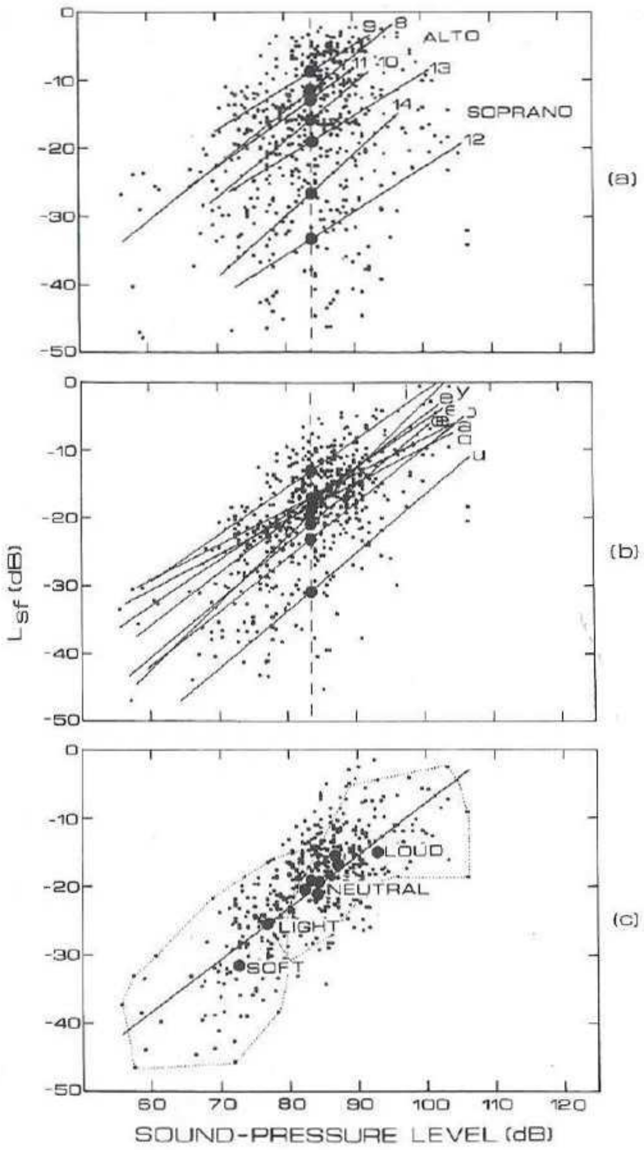

Fig. 8.1. Example of the analysis of the level of the singer's formant for 567 vowels, sung by seven female singers at F_0 = 220 Hz.
(a) Original data and the regression line between L_{sf} and overall SPL for each singer (8-10, alto; 11-12, mezzo-soprano; 13-14, soprano).
(b) Singer-normalized data and the regression line between L_{sf} and overall SPL for each vowel.
(c) Data, normalized for singers and vowels, plus the regression line. The positions of modes of singing are indicated. For the modes of singing soft and loud the contour of data points has been drawn (dotted line).

8.3 Results

Using the procedure described above all vowel sets, with F_0 varying between 98 and 880 Hz, were analyzed. We first obtained the singer-specific L_{sf} and the vowel-specific L_{sf}. For the F_0 ranges of 98 up to 392 Hz (modal register) for males and F_0 = 220 and 392 Hz for females, the effect of F_0 was insignificant for both the singer-specific and the vowel-specific L_{sf}. Therefore we give in Fig. 8.2 representative data of both quantities for F_0 = 220 Hz. A remarkable difference between males and females was revealed: whereas there were no substantial differences in L_{sf} for male singers ($-15>L_{sf}>-19$ dB), for female singers L_{sf} varied between

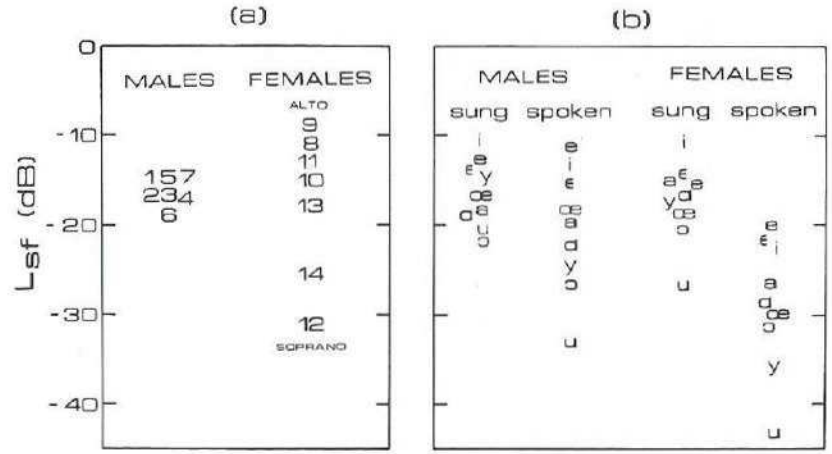

Fig. 8.2. Singer-specific L_{sf} (a) and vowel-specific L_{sf} (b) for F_0 = 220 Hz. Data for spoken vowels were adapted from Klein et al. (1970) for males and from van Nierop et al. (1973) for females. The numbers of the singers are the same as in Table 2.1.

-8 and -32 dB, and this variation was related to voice classification. Alto singers 8, 9, and 10 typically have the highest sound level of the singer's formant, higher even than for male singers, while (mezzo)soprano singers may show a singer's formant more than 20 dB lower in sound level than alto singers. As was mentioned before, these results hold for wide F_0 ranges. There were no great differences in the vowel-specific L_{sf} between males and females except the lower value for the vowel /u/ for female singers.

Additionally, we were able to make an interesting comparison with sound level data for spoken vowels (from the same frequency band) obtained by Klein et al. (1970) for 50 Dutch males and by van Nierop et al.(1973) for 25 Dutch females. Their data are also presented in Fig. 8.2. It may seem counter-intuitive to speak of a "singer's formant" for spoken vowels. However, within our simple and straightforward definition of the singer's formant as the sound level in a particular frequency band, there is no essential problem about doing so at this stage; but see the discussion in Section 5.4.

For males most sung and spoken vowels did not differ much with respect to L_{sf}, except the vowels /y/, /ɔ/, and /u/, which had considerably lower L_{sf} values for spoken vowels. For females there was a systematic difference between spoken and sung vowels in L_{sf} for most vowels, but relatively larger for the vowels /y/ and /u/. For a possible explanation of these differences between spoken and sung vowels we have to consider the relation between L_{sf} and overall SPL. As can be seen in Fig. 8.1b, the slope of the regression line between L_{sf} and overall SPL is not the same for all vowels: for some vowels L_{sf} rises more quickly with increasing overall SPL than for other vowels. A quantification of this effect can be given as the increase in dB of L_{sf} when overall SPL rises by 10 dB. These data are presented in Table 8.1 for all vowels, as the average values over male modal phonations (F_0 = 98, 131, 220, 392 Hz), male falsetto phonations (F_0 = 392 Hz), female lower F_0 range (F_0 = 220, 392 Hz), and female higher F_0 range (F_0 = 659, 880 Hz), separately. Largest differences among vowels were found in the male modal register. In absolute sound level the singer's formant for /u/ rises by 10+15.4=25.4 dB in the modal register when overall SPL rises by 10 dB, which is 10 dB more than for the vowel /ɛ/. It can be seen that for both male singers (modal register) and female singers (low F_0 range) the vowels /ɔ/, /y/, and /u/ show the greatest increase in L_{sf}. On the average the increase is less in the male falsetto register than in male modal register; no important differences among vowels were found for the

Table 8.1. Increase of L_{sf} in dB for nine vowels when overall SPL increases by 10 dB, obtained from singer-normalized data. Average results for a number of F_0 values are given.

	Males		Females	
	modal	falsetto	low F_0	high F_0
F_0 (Hz)	98,131,220,392	392	220,392	659,880
ɛ	5.8	2.1	7.0	5.5
a	6.9	2.6	5.7	5.0
ɑ	7.2	4.0	5.8	6.5
e	6.6	5.0	8.4	6.0
œ	8.3	9.3	7.9	5.5
i	8.7	7.4	8.4	5.0
ɔ	9.2	6.1	8.3	6.2
y	12.9	7.6	11.6	5.9
u	15.4	10.1	10.7	6.7

female high F_0 range.

On the basis of these results we may explain the differences in L_{sf} for spoken vowels (Fig. 8.2). We predict what would happen to the L_{sf} of sung vowels if overall SPL decreased by about 7 dB for males and about by 15 dB for females. The corresponding decrease in L_{sf} can be derived from the data in Table 8.1 in the columns "male modal register" and "female low F_0", respectively. For the vowel /u/ (males) this results, for example, in a value of -20.5-0.7x15.4=-31.3 dB for L_{sf}; for the vowel /i/ (males) L_{sf} becomes -11.0-0.7x8.4=-16.9 dB. These values, and those for all other vowels, agree fairly well with the data for spoken vowels. Although no data of overall SPL were available for the spoken vowels, the suggested difference with sung vowels of 7 dB for males and of 15 dB for females seems plausible, although possibly somewhat large for females. Thus, the differences in L_{sf} between spoken and sung vowels may largely be explained as the effect of a difference in overall SPL.

After the normalization of L_{sf} for differences due to singers and vowels, regression lines of L_{sf} as a function of overall SPL were computed for all subsets, in the same way as demonstrated in Fig. 8.1c, and these regression lines are presented in Fig. 8.3a. It can be seen that L_{sf} varies from -35 to -3 dB for all F_0 values, except for male falsetto register and the female high F_0 range, which occupy lower values of L_{sf}. The differential effect of fundamental frequency on L_{sf} is that the same value of L_{sf} can be reached for a higher F_0 only when overall SPL is increased. In other words, when overall SPL is kept constant, L_{sf} decreases when F_0 increases

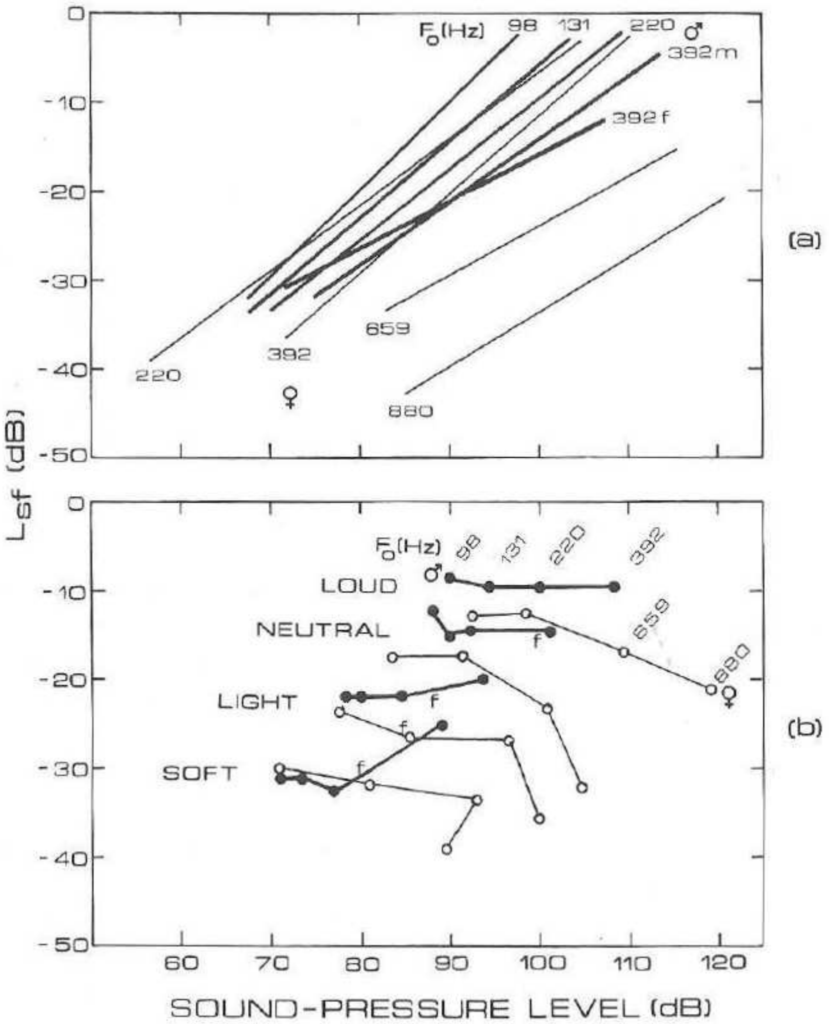

Fig. 8.3. (a) Regression lines between L_{sf} and overall SPL for data, normalized for singers and vowels, as a function of F_o, for males and females separately. (b) L_{sf} of modes of singing as a function of overall SPL and F_o.

(Schultz-Coulon et al., 1979).

In Fig. 8.3b the average L_{sf} for different modes of singing is plotted as a function of F_0 and overall SPL. For the sake of clarity only data points for the modes soft, light, neutral, and loud have been given. Data for the other modes compare well with the neutral mode. Separately for males and females, for each mode of singing the data points of increasing F_0 values are connected. It can be seen that in the male modal register L_{sf} is approximately constant over the four modes of singing due to a balanced increase of overall SPL and F_0. This indicates that with increasing F_0, vocal effort (and overall SPL) is adapted to each mode in such a way that L_{sf} is kept constant, which probably leads to or is even governed by a perceptually constant quality of the voice. Only in the soft mode (F_0 = 392 Hz, males) did this adaptation fail because overall SPL rose too much, or could not be kept at a sufficiently level. For the male falsetto phonations L_{sf} has, compared to male modal register (F_0 = 392 Hz), a 5 dB lower average value for the modes loud, neutral, and light a 5 dB lower value, but the average overall SPL of the falsetto phonations was also 9 dB. For female singers L_{sf} is constant over each mode of singing for F_0 = 220 and 392 Hz only; for the loud, neutral, and light modes it is at a 3 dB lower level than for male singers, and for the soft mode it is at the same level. Note, however, that for the loud, neutral, and light modes the difference in L_{sf} between the modes for female singers is similar to the difference between modes for male singers. At F_0 = 659 Hz, for the modes neutral and loud, L_{sf} is decreased; at F_0 = 880 Hz L_{sf} is decreased considerably for all modes of singing.

8.4 Discussion

Our study of the sound level of the singer's formant has revealed several remarkable characteristics, which could be obtained by a systematic analysis of a very large number of profesionally sung vowels. First, Fig. 8.2 showed that male voice classification had not much influence on the level of the singer's formant whereas reverse was true for females. This result is in agreement with the spectral characteristics of voice classification, given in Chapter 5. It is surprising that at the same F_0, some alto singers even showed a stronger singer's formant than bass singers. The singer's formant has often been demonstrated for low male voice classifications and it has been argued that especially for these voices the singer's formant has perceptual relevance, making a singer stand out from orchestral accompaniment when F_0 is low (Sundberg, 1974). Although our results do not contradict the possibility of this perceptual relevance, the reasoning does not uniquely apply to bass singers. It can only be said that it does not hold for some soprano singers.

The average sound level of the singer's formant was, of course, vowel dependent. However, in accordance with the general results on the spectral effect of increasing overall SPL (Chapter 7), the effect of vocal effort on L_{sf} was highly vowel dependent too. This may explain to some extent the difference between L_{sf} for sung and spoken vowels, since sung vowels are on the average much louder. Some possible glottal mechanisms underlying the SPL dependence of L_{sf} have been discussed already in Sections 5.3 and 6.3.

The results on the relation between F_0, vocal intensity, and L_{sf} (Fig. 8.3a) confirm the findings of Schulz-Coulon (1979) that L_{sf} increases for increasing vocal intensity (F_0 constant) and decreases for increasing F_0 (vocal intensity constant). Above F_0 = 392 Hz the decrease of L_{sf} is considerable, as was also demonstrated by the grand-average spectra for each F_0 value, presented earlier in Chapter 4.

The effect of mode of singing on L_{sf} is most interesting. It shows that a singer maintains the same L_{sf} for a certain mode of singing by increasing overall SPL when F_0 is raised. We may speculate that this leads to a highly stable perceptual quality of a vowel.

It can be seen in Fig. 8.3b that for the modes of singing light, neutral, and loud, L_{sf} is on the average about 4 dB lower for female singers than for male singers. This shows that in the same mode of singing the singer's formant is, on the average, less prominent for female singing, which seems to agree with an often mentioned (i.e., Bartholemew, 1934), but not experimentally verified perceptual observation. The same result applies to the male falsetto register relative to male modal register.

The sound-level range of the singer's formant was, for the various factors: (1) male voice classification: 4 dB (Fig. 8.2a), (2) female voice classification: 24 dB (Fig. 8.2a), (3) vowels: 16 dB (Fig. 8.2b), (4) overall SPL: more than 30 dB (Fig. 8.3a); between the modes of singing soft and loud 23 dB (Fig. 8.3b), and (5) F_0, up to 392 Hz: between 9 and 14 dB (Fig. 8.3a). The combination of all these factors introduces such a great variation in L_{sf} that a more restricted definition of the singer's formant as a perceptual phenomenon, the "ring", seems to be required. Lacking perceptual data, we may hypothesize that such a definition should at least require a minimum sound level of L_{sf}, independent of F_0. This is suggested by the stable level of L_{sf} over different modes of singing, which probably has a perceptual origin. If the singer's formant may be assumed to be present in male and female neutral and loud singing and not in the light and soft modes of singing, its perceptual threshold may be found around L_{sf} = -20 dB. In that case most male and female singers, as well as most sung vowels, exceed on the average this threshold. Exceptions are the soprano singers, the sung vowel /u/, male spoken back vowels, all female spoken vowels, F_0 values above 392 Hz, and the modes of singing light and soft (by definition).

In addition, it is probable that a simple sound level threshold is insufficient to define the singer's formant as a perceptual phenomenon. Pressed singing, for instance, may satisfy this sound level condition. Therefore, it is likely that the singer's formant should be embedded in a high quality phonation type, such as "flow" phonation (Sundberg and Gauffin, 1979). In that case spectral conditions are not only limited to the frequency range between 2 - 4 kHz; especially the relative sound level of the fundamental may be of importance too: this relative level should exceed a certain threshold. As a consequence the presence of the singer's formant will be limited still more, probably to singers only.

PART II

THE TIMBRE OF SUNG VOWELS

9 INTRODUCTION

In Chapter 1 we introduced timbre as "that attribute of auditory sensation in terms of which listeners can judge that sounds having the same pitch and loudness are dissimilar" (ASA, 1960). The importance of timbre in singing was outlined and it was concluded that there was a need to investigate (1) the representation of the timbre of stationary sung vowels, (2) the physical prediction of this representation on the basis of the 1/3-oct spectra of the vowels, (3) the relation between the timbre of sung vowels and its verbal attributes, and (4) the acoustical correlates of verbal attributes of timbre. The necessary studies, presented in the following chapters, include both listening experiments and spectral analyses. In this chapter we briefly introduce the level of auditory perception on which our listening experiments focus and present some general characteristics of the two types of listening experiments we performed. The spectral analyses of the vowel stimuli follow the general procedures given in Chapter 4.

Plomp (1970) pointed out that timbre is a multidimensional attribute of complex stationary harmonic tones. These sounds can be represented as points in a multidimensional perceptual space in which distance corresponds to dissimilarity in timbre: the larger the distance in this space, the more the sounds are perceptually dissimilar in timbre. Since such a representation of timbre is only based on perceived dissimilarity of complex sounds, it has also been described as a psycho-acoustic representation, which has its origin in properties of the peripheral auditory level of perception. For spoken or sung vowels a specific quality can be assigned to a phonemic identity, which process is thought to take place at a higher, phonetic, level of perception. This phonetic level is, for instance, sensitive to formant frequencies but rather insensitive to spectral slope (Klatt, 1982). Because in singing the phonemic identity of a vowel is often secondary to timbre, we have limited ourselves in the present study to the psycho-acoustic representation of the timbre of stationary sung vowels.

We distinguish two types of perception experiments on the basis of which we can study the psycho-acoustic representation of timbre. These are the so-called "non-verbal" and "verbal" listening experiments. In the non-verbal listening experiment the listener is requested to compare two pairs of sounds which have one sound in common. The listener has to judge which pair is most dissimilar in timbre. Thus, the listener does not need to specify verbally the qualities of the sounds presented. This is advantageous because such an experiment can be performed by unexperienced listeners. The results of a non-verbal listening experiment make it possible to compute a representation of the timbres of the stimulus sounds as points in a multidimensional space.

In contrast, in the verbal listening experiment, we investigate timbre with the aid of terms which are associated with specific characteristics of timbre. In this case a listener is requested to compare pairs of sounds and to judge for each pair, for instance, which of the two sounds was the more "sharp", and so on for 21 different adjectives. The verbal listening experiment again results in a representation of the timbres of sounds as points in a perceptual space. It will be shown that the verbal perceptual configuration of sounds obtained in this way is similar to the one obtained in non-verbal experiments. In addition, the verbal experiments make it possible to assign verbal terms, such as sharp, dark, light, and metallic, to regions of the perceptual space. Because this assignment is possible for each listener individually, we could investigate whether listeners agree about the interpretation of the terms.

The psycho-acoustic differences in timbre can be predicted from the differences in the spectral envelopes of sounds. This has been shown by Pols et al.(1969) for spoken vowels and by Plomp (1970) for musical sounds. The prediction is based on the difference in sound level (in dB) in 1/3-oct bandpass filters, which correspond to the concept of the critical band in human hearing. As was concluded already in Chapter 1, more sophisticated models of the peripheral auditory process (e.g., Bladon and Lindblom, 1980) so far have not resulted in much better predictions. The correspondence between the psycho-acoustic representation of vowel sounds in a perceptual space and the spectral representation based on 1/3-oct spectra in a spectrum space makes it possible to study timbre differences acoustically, especially those which are typically related to descriptive terms of timbre.

The present study is focused on the description of distinctive timbre differences between advanced student singers. These differences will be studied by means of non-verbal listening experiments (Chapter 11) and verbal listening experiments (Chapter 12) for the vowels /a/, /i/, and /u/, sung at fundamental frequencies of 131, 220, and 392 Hz. It should be emphasized that this study is based on different singing material than that presented in Part I.

10 MATERIAL

In a concert hall recordings were made of vowels, sung by nine female and eight male singers. All but one (professional bass-baritone singer 1) were advanced students of the Sweelinck Conservatory in Amsterdam, aged between 19 and 26, with 3 to 7 years of vocal training. The microphone distance was 0.3 m, so that the direct sound predominated. According to their own voice classifications, the group consisted of: 2 bass-baritone, 2 baritone, 4 tenor, 2 alto, 3 mezzo-soprano, and 4 soprano singers (Table 10.1). The vowels /a/, /i/, and /u/ were sung at a comfortable level at fundamental frequencies (F_0) of 131 (C3), 220 (A3), and 392 Hz (G4). Some male singers performed at F_0 = 392 Hz in modal register and falsetto register as well.

Since we only wanted to investigate spectral attributes of timbre, temporal variations such as vibrato were removed from the vowel sounds. To accomplish this, each vowel sound was digitized (10 kHz sampling frequency); subsequently a single period with a fixed number of samples was segmented from the central part of the vowel, and this period was repeated to obtain a stimulus duration of 400 ms. To avoid clicks at the onset and offset of each stimulus, the sound level increased and decreased smoothly during the first and last 40 ms, respectively.

Ten different subsets of eight or nine vowel sounds were made by combining phonations of the same vowel with the same F_0 sung by different singers. The subsets varied according to vowel type (/a/, /i/, and /u/, subsets II, IV, V (males) and subsets VIII, IX, X (females)), and according

Table 10.1. Classification of the singers

male singers	female singers
1 bass-baritone	9 alto
2 bass-baritone	10 alto
3 baritone	11 mezzo soprano
4 baritone	12 mezzo soprano
5 tenor	13 mezzo soprano
6 tenor	14 soprano
7 tenor	15 soprano
8 tenor	16 soprano
	17 soprano

Table 10.2. Vowel, fundamental frequency and participating singers (labeled according to Table 10.1) for the eleven subsets. f indicates falsetto register, t indicates a tenor-like register produced by baritone singer 3.

Subset	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
Vowel	a	a	a	i	u	a	a	a	i	u	'Halleluja'
F_0(Hz)	131	220	392	220	220	220	220	392	392	392	-
Singers	1	1	1	1	1	1	9	9	9	9	1
	2	2	3	2	2	5	10	10	10	10	2
	3	3	3t	3	3	6	11	11	11	11	3
	4	4	6	4	4	8	12	12	12	12	4
	5	5	8	5	5	9	13	13	13	13	5
	6	6	1f	6	6	10	14	14	14	14	6
	7	7	3f	7	7	11	15	15	15	15	7
	8	8	7f	8	8	13	16	16	16	16	8
			8f			15	17	17	17	17	

to fundamental frequency (vowel /a/ for F_0 values of 131, 220, and 392 Hz, subsets I, II, III (males) and subsets VII, VIII (females)), see Table 10.2. Subset III combined phonations sung in the modal register and the falsetto register by male singers (F_0 = 392 Hz). Subset VI combined a selection of vowels /a/ by both male and female singers (F_0 = 220 Hz). Because of the limitations imposed by the listening experiments, the maximum number of stimuli in each subset was nine. The loudness of the vowels in each subset was equalized by means of a subjective matching procedure.

Recordings were also made of a Dutch folk song sung by the eight male singers. From this folk song the phrase 'Halleluja' (Fig. 10.1) was segmented. Since the singers were free in their interpretation of this song, their recordings varied in time between 2.6 and 4.4 s. The phrase was used in the verbal listening experiment to investigate the description of timbre in a real song in contrast to the description of timbre of the "electronically" sounding stationary vowels. The loudness of the phrases was equalized subjectively by means of a subjective matching procedure; the resulting phrases formed subset XI. No comparable subset of female singers was made because of the very time-consuming listening experiments.

Fig. 10.1. Musical notation of the song phrase 'Halleluja' (subset XI).

11 NON-VERBAL LISTENING EXPERIMENTS

The perception of timbre differences between vowels, sung at fundamental frequencies of 131, 220, and 392 Hz by 17 singers with voice classifications ranging from bass-baritone to soprano, was investigated in a non-verbal way. Both musicians and non-musicians participated as listeners. Using INDSCAL analysis we derived (1) vowel configurations in a multidimensional perceptual space, and (2) individual listener weightings of perceptual dimensions. The perceptual vowel configurations could be predicted on the basis of differences in 1/3-octave spectra between the vowels. No substantial differences were found between the judgments of musicians and non-musicians.

Part of a paper with R.Plomp, submitted to J. Acoust. Soc. Am.

11.1 Introduction

In the non-verbal listening experiment a listener was requested to judge the perceptual dissimilarity of pairs of stationary vowel sounds of equal loudness and pitch without specifying the kind of dissimilarity. Comparable experiments have been performed by Pols et al. (1969) for the rather large timbre differences between spoken vowels, at a low F_0 value. The present experiment was designed (1) to test the validity of predicting timbre dissimilarities from spectral differences for the case of the relatively small interindividual timbre differences in one and the same vowel sung by different singers, (2) to test the validity for fundamental frequencies up to 392 Hz (for higher F_0 values the timbre differences in the same vowel due to differences between singers become marginal), (3) to test the hypothesis that musical training has no significant effect in this type of listening experiment, and (4) to evaluate the reliability of the experiment.

11.2 Procedure

11.2.1 Listeners

We used two categories of listeners. The first category consisted of seven non-musicians (who had never had any musical training), the second category of nine musicians (five singers and four singing teachers). In most non-verbal listening experiments non-musicians were involved, as it was hypothesized that the psycho-acoustic representation of sounds depends on fundamental properties of the auditory periphery rather than on the degree of musical training. This hypothesis was verified for subset V, which was also judged by nine musicians. All listeners had normal audiograms. They were paid for their services.

11.2.2 Method

To map the psycho-acoustic representation of timbre differences we used the method of paired comparison of pairs. This method is a modification of the triadic comparison technique, used, for instance, by Pols et al. (1969) and Plomp (1970). A triadic comparison between stimuli A, B, and C, is split up in three paired comparisons of pairs with a common first stimulus: AB vs. AC, BA vs. BC, and CA vs. CB. In contrast to the triadic comparison, in which a listener can listen repeatedly to the three stimuli before deciding which pair is most similar and which is most dissimilar, in the present procedure each pair of pairs is presented only once, after which the listener has to indicate which pair contained the more similar stimuli. This technique reduces experimentation time and makes the experiment uniform for all listeners. The listener, seated in a sound proof room, heard the stimuli monaurally at a comfortable level through headphones (Beyer DT-48). All possible pairs of pairs of vowels in a subset were presented in random order, while each vowel of a subset was presented equally frequently in the first and second pairs. To eliminate order effects, half of the listeners heard the pairs in reversed order (AC-AB instead of AB-AC). To test reliability, subset V was judged in test and retest. In the retest case the order of presentation of pairs was reversed. The experiment was computer controlled, which meant that stimulus generation, timing, and response processing were handled by the computer (PDP 11/10).

11.2.3 INDSCAL analysis

The results of the paired comparison experiment were converted for each listener into a dissimilarity matrix. Every time a particular pair of vowel stimuli was judged as more similar, it scored one point. The total number of points which could be assigned to a pair could vary between zero and the total number of vowel stimuli minus one. The dissimilarity matrices were analyzed by means of a multi-dimensional scaling technique (INDSCAL, quasi non-metric version, Carroll and Chang, 1970). This analysis results in a multidimensional space (object space) in which the vowels of a subset are represented in such a way that their distances correspond as closely as possible to their subjective dissimilarities. Different listener judgments of perceptual relations between vowels are accounted for by individual weighting factors for each dimension. These factors are presented in a subject space. The dimensionality of the INDSCAL solution was chosen on the basis of the results of matching the configuration in the object space with the spectral representation of the stimuli, discussed in the next section. The correlation in each matched dimension had to be significant beyond the 0.05 level.

11.2.4 Matching of the perceptual and spectral vowel configurations

For each subset of stimuli we had available both a perceptual and a spectral configuration of the vowels. To investigate the agreement between these configurations we matched them, using the procedure of rotation to maximal congruence (Schönemann and Carroll, 1970) between the spectral and the perceptual configurations. As a measure of fit we computed (1) the correlation between vowel coordinates in each perceptual and matching spectral dimension and (2) the coefficient of alination S (Lingoes and Schönemann, 1974). S varies between 0 (perfect fit) and 1 (unrelated configurations). The perceptual (object) space derived by INDSCAL is normalized (equal variance in all dimensions). We determined, by an iterative procedure, weighting factors for the dimensions of this perceptual space, which optimized the fit measure S. The correlation coefficients per dimension were practically uninfluenced by these weighting factors.

After having obtained a successful matching between vowel configurations in the perceptual and the spectrum spaces, we are now able to interpret the perceptual dimensions found in spectral terms. We will not consider this for the non-verbal experiment because we are not convinced that in this case the dimensions found by INDSCAL have a unique perceptual relevance. Moreover, the results of the verbal experiment (Chapter 12) give a direct spectral interpretation of the meaningful perceptual dimensions.

11.3 Results and discussion

An example of a perceptual (object) space, the matched spectrum space and the listener (subject) space is shown in Fig. 11.1 for subset V (/u/, sung by eight male singers at F_0 = 220 Hz). This subset was judged by nine singers and singing teachers (open squares) and seven non-musicians (filled squares). Five listeners from the latter group judged in test and retest. The upper panels of figure 11.1 illustrate the very good agreement between the vowel configuration in the three-dimensional perceptual space (filled circles) and the matched configuration from the spectrum space (open circles). It can be seen from the subject space (lower panels) that the interindividual differences for the musicians are about as large as those for non-musicians. Separate analyses of the data from musicians and non-musicians showed similar object spaces (not shown). These results support the view that in this kind of non-verbal psycho-acoustical experi-

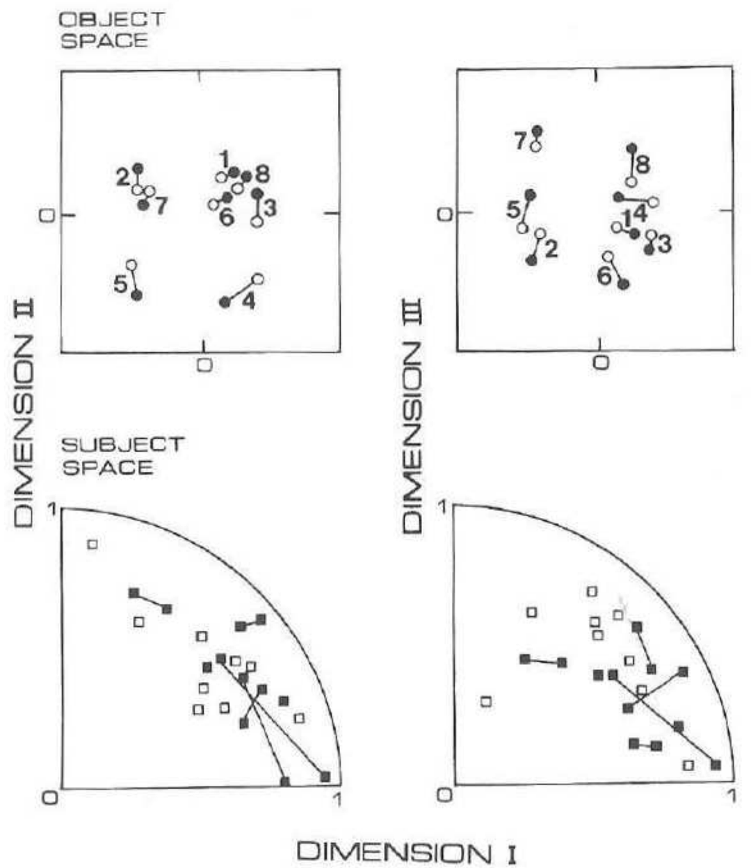

Fig. 11.1. Result of an INDSCAL analysis on data from non-verbal judgments of the vowel /u/, sung by eight male singers at F_0 = 220 Hz. The upper panels show the I-II and the I-III plane of the object space and spectrum space combined. Filled circles form the vowel configuration obtained from INDSCAL, open circles present the best fitting spectral vowel configuration. The lower panels show the corresponding planes from the subject space of INDSCAL. Open squares are musicians, filled squares are non-musicians. Positions of non-musicians who judged in test and retest are connected.

ments the musical training of listeners does not play a part.

Fig. 11.1 (subject space) shows that the differences between test and retest for five non-musicians were nonnegligible for two listeners; the result of their retest was highly one-dimensional (dimension I). Also highly one-dimensional, but for dimension II, was the result of a musician. In these cases, listeners seem to concentrate on only one spectral aspect of timbre difference and judge all stimuli accordingly. That this occurred in two out of the six retest judgments may possibly indicate a concentration on perceptual main effects after repeated presentations. This behavior, however, did not prevail for most listeners in their first judgment of a stimulus set.

A summary of data on the spectral variance in each of the ten vowel subsets and the result of matching the perceptual vowel configuration with the vowel configuration in the spectrum space is presented in Table 11.1. The total spectral variance in our stimuli varied between 103 and 332 dB^2, due to differences between singers. This range corresponds well with the spectral variance associated with the factor "singers" and the interaction between "singers" and "vowels", equal to about 180 dB^2, found in Chapter 3 for 14 professional singers. The total spectral variance in a subset depended on the value of F_0 and the vowel. Highest values were obtained for

Table 11.1 Results of matching the vowel configurations in the perceptual space, derived by INDSCAL from non-verbal listening experiments, and the spectrum space. Each subset was judged by seven listeners, except subset V which was judged by 21 listeners (see text and Fig. 11.1). S represents the coefficient of alination, D is dimension.

subset			vowel	F_0 (Hz)	total spectral variance (dB^2)	percentage of total spectral variance D1	D2	D3	D4	Σ	correlation between perceptual and spectrum coordinates along dimensions D1	D2	D3	D4	S
I	8	M	/a/	131	211	34	21	19	17	91	0.92	0.95	0.97	0.95	0.39
II	8	M	/a/	220	158	36	32	14	7	89	0.95	0.96	0.87	0.82	0.44
III	9	M	/a/	392	332	38	37	11	-	86	0.92	0.92	0.72	-	0.53
IV	8	M	/i/	220	239	37	22	16	14	89	0.91	0.87	0.83	0.94	0.51
V	8	M	/u/	220	255	45	21	19	-	85	0.95	0.94	0.92	-	0.47
VI	9	M+F	/a/	220	200	30	26	26	7	89	0.94	0.96	0.86	0.88	0.48
VII	9	F	/a/	220	114	48	21	-	-	69	0.91	0.77	-	-	0.68
VIII	9	F	/a/	392	103	34	26	20	-	80	0.77	0.66	0.92	-	0.71
IX	9	F	/i/	392	180	28	24	19	18	89	0.96	0.95	0.93	0.93	0.42
X	9	F	/u/	392	248	38	36	15	-	89	0.96	0.89	0.74	-	0.52

low F_0 values, while total spectral variance increased for vowels in the order /a/, /i/, and /u/. The spectral differences between the modal register and the falsetto register introduced the largest spectral variance (subset III).

The number of dimensions for which the correlations between vowel coordinates along the perceptual and matched spectral dimensions were significant beyond the 0.05 level varied between two and five. This number is probably related to the amount of spectral variance present in a dimension. It was found that the least significant dimension explained on the average 31 dB^2 of spectral variance. This corresponds to a standard deviation of, roughly, between 1 and 2 dB for each frequency band (depending on F_0) if spectral variance is uniformly distributed over frequency bands. If spectral variance is concentrated in a single frequency band, this would correspond to a standard deviation of 5.5 dB in that band. De Bruyn (1978) concluded from investigations on timbre dissimilarity of complex tones that two complex tones are distinguished well by listeners for a mean difference in sound levels of between 3 and 5 dB in each 1/3-oct band. The difference limen for individual harmonics in vowel sounds was estimated by Kakusho et al. (1971) to be less than 2 dB for most vowels. These thresholds roughly indicate that in our investigation the correspondence between spectral representation of timbre and psycho-acoustic representation is valid up to the perceptual threshold of timbre differences. This limit determined the dimensionality of the perceptual vowel configuration for all F_0 values investigated.

The high correlation coefficients in Table 11.1 indicate that the representation of vowels in a spectrum space gives a reliable description of the small psycho-acoustic timbre differences investigated, equally well for all F_0 values investigated up to 392 Hz. For low F_0 this is in agreement with results obtained by Nord and Sventelius (1979) concerning just-noticeable differences in formant frequency, and by Klatt (1982) for a number of physical manipulations of a single vowel.

We may conclude, therefore, that (1) the approach is suitable for the description of small timbre differences in the same vowel sung by different singers, (2) the approach is suitable up to high F_0 values, and (3) there is no difference in judgments by musicians and non-musicians, whereas for both groups the interindividual differences were fairly large.

12 VERBAL LISTENING EXPERIMENTS

Timbre differences between vowels and song phrases were investigated on the basis of judgments by musicians and non-musicians, using 21 semantic bipolar scales. Using MDPREF analysis we derived (1) vowel configurations in a multidimensional space, and (2) directions in this space associated with semantic scales, separately for each listener. The perceptual vowel configurations were comparable with those obtained from the non-verbal listening experiments (Chapter 11). Semantic scales clustered into the categories singing technique, general evaluation, temporal aspects, clearness, and sharpness. These five categories were not independent and could be described in two dimensions. Non-musicians judged most scales in one manner according to sharpness. Musicians differentiated better in the use of semantic scales, especially for the song phrases. Acoustically, only sharpness had a consistent interpretation for the various stimulus sets and was roughly related to the slope of the spectrum. Other perceptual dimensions had a vowel-dependent acoustical interpretation; there was little agreement among listeners in the verbal description of these dimensions.

Part of a paper with R.Plomp, submitted to J. Acoust. Soc. Am.

12.1 Introduction

Many terms are used for the verbal description of the timbre of sung vowels. As has been said in Chapter 1, these terms play an important role in voice pedagogy, but agreement about their interpretation is often absent. The difficulty of describing timbre is not restricted to voice pedagogy. Stumpf (1890) already compared the wealth of verbal timbre characteristics with descriptions of the quality of wine, used by merchants, while Sonninen and Damsté (1971) recall the urgent need for a well-defined terminology in the field of logopedics and phoniatrics. On the other hand, it should be realized that the vocabulary of voice pedagogy not only refers to a specific acoustical attribute of the voice but often also to the associated singing technique (e.g. covered, open, throaty, pressed, free), as was pointed out by Van den Berg and Vennard (1959). For the present experiment we chose out of the very large number of available descriptive adjectives, 21 pairs with opposite meaning, such as light-dark. A pair of adjectives defines a semantic scale along which the timbre of sounds can be ordered, for instance, from very light to very dark.

Verbal listening experiments with these 21 semantic scales were performed on five of the ten subsets of stationary vowels used in the non-verbal listening experiment and on the stimulus subset of song phrases. The latter subset was included to investigate the influence of dynamic acoustical variation on timbre. We investigated whether for a vowel subset the verbal listening experiment led to the same multidimensional configuration of stimulus points as the non-verbal experiment (Section 12.3.4). If this is the case it may be concluded that verbal listening experiments also describe timbre differences at the peripheral auditory level of perception.

In the non-verbal listening experiment (Chapter 11) we showed that the number of perceptual dimensions of timbre variation in a single vowel for which a significant correlation to spectrum space dimension exists is restricted to between two and four. On the basis of this small number of perceptual dimensions we may expect that many verbal attributes of timbre are not independent. This view is supported by the results of a factorial analysis of verbal attributes of timbre by Von Bismarck (1974a), who found, for complex stationary harmonic tones, only one prominent attribute: sharpness. We investigated relations between the 21 semantic scales for two groups of listeners: musicians and non-musicians (Section 12.3.2). Furthermore, the agreement between individual listeners upon the interpretation of

semantic scales in relation to the stimuli is discussed in Section 12.3.3.

A verbal attribute of timbre, whether it has a general or an individual meaning, should refer to specific acoustical qualities. Von Bismarck (1974b) and Benedini (1980) demonstrated that sharpness was related to the relative importance of higher harmonics. On the basis of an excellent match of vowel configurations in the perceptual space and the spectrum space in the present experiment, we could study the acoustical interpretation of semantic scales for each individual listener. This was possible for the five subsets of stationary vowels (Section 12.3.5), and an attempt was made for the acoustically more complex subset of song phrases (Section 12.3.6).

12.2 Procedure

12.2.1 Listeners

As was the case in the non-verbal listening experiment, both musicians and non-musicians participated as listeners. As will be pointed out in Section 12.2.3 not all listeners judged all stimulus subsets. The listeners were paid for their services.

12.2.2 Semantic bipolar scales

For the verbal listening experiments semantic bipolar scales were used. Each semantic scale consisted of two adjectives with opposite meanings, describing timbre characteristics such as light-dark and colorful-colorless. For the determination of the set of semantic scales to be used in our listening experiment we first collected 50 scales from related studies on timbre (Isshiki et al., 1969; Donovan, 1970; von Bismarck, 1974a; Fagel et al.(1983); Boves, 1984) and from the literature on singing (Vennard, 1969; Hussler, 1976). These semantic scales were rated by seven experts (speech therapists, singing teachers) on their suitability for describing the timbre of sung vowels. Of the 50 scales, 21 were generally judged to be suitable, see Table 12.1. Of these, scales 1 to 14 and scale 21 were regarded as commonly known adjectives for the description of timbre and were used by all listeners. The scale vibrato-straight (21) was not used for stationary vowel sounds but only for judgments of the song phrase. The scales free-pressed (15), open-throaty (16), and open-covered (17) were considered to be evaluative of singing technique. The scales dramatic-lyric (18), soprano-alto (19), and tenor-bass (20) were intended to investigate relations between timbre and voice classification. These six scales were used only by musicians.

12.2.3 Method

The most common method using semantic scales is the semantic differential, in which a listener has to rate each stimulus on, for example, a 7-point bipolar scale. Category 1 and category 7 describe the opposite adjectives of the semantic scale. Due to variability in rating among listeners and to the procedure of analysis (factor analysis), in which the results of all listeners are averaged, this method requires a large number of listeners. No information about individual judgments is left in the final result. Since we did not have a large number of musicians at our disposal and, more importantly, we were particularly interested in interindividual differences in the interpretation of adjectives, we followed a different procedure. This procedure consisted of a paired comparison of all

Table 12.1. Bipolar semantic scales used in the verbal listening experiment. Target adjectives are underlined.

1.	light	-	heavy	(licht	-	zwaar)
2.	light	-	dark	(licht	-	donker)
3.	sharp	-	dull	(scherp	-	dof)
4.	clear	-	dull	(helder	-	dof)
5.	full	-	thin	(vol	-	ijl)
6.	shrill	-	deep	(schel	-	diep)
7.	high	-	low	(hoog	-	laag)
8.	rough	-	smooth	(ruig	-	vlak)
9.	angular	-	round	(hoekig	-	rond)
10.	strong	-	weak	(krachtig	-	zwak)
11.	cold	-	warm	(koud	-	warm)
12.	colorful	-	colorless	(kleurrijk	-	kleurloos)
13.	metallic	-	velvety	(metaalachtig	-	fluweelachtig)
14.	melodious	-	unmelodious	(welluidend	-	onwelluidend)
15.	free	-	pressed	(vrij	-	geknepen)
16.	open	-	throaty	(open	-	kelig)
17.	open	-	covered	(open	-	gedekt)
18.	dramatic	-	lyrical	(dramatisch	-	lyrisch)
19.	soprano	-	alto	(sopraan	-	alt)
20.	tenor	-	bass	(tenor	-	bas)
21.	vibrato	-	straight	(vibrato	-	strak)

stimuli of a subset on a semantic scale. The listener had to judge which of the two stimuli presented was closer to a given target adjective of a semantic scale, for example: which of two stimuli was darker (on the semantic scale light-dark). The chosen stimulus scored one point. After all possible pairs of stimuli had been judged, the total number of points obtained by each stimulus was used to rank the stimuli from light to dark. In Table 12.1 the adjectives used as targets are underlined.

Experiments performed in this way are very time consuming. Hence we reduced the number of subsets to five subsets of stationary vowels (III, IV, V, VI, VIII) and one subset of song phrases (XI). Three half-day sessions were planned for each listener to complete all measurements. Most of the musicians could not complete all measurements in the time available; therefore the number of listeners per subset varied. The stimuli were presented in random order but equally frequently in first and last position of a pair. Half of the total number of listeners heard the stimuli of a pair in reversed order. The experiments were computer controlled.

12.3 Results

12.3.1 Reliability

To judge the reliability of the results of a listener in a paired comparison experiment on a single semantic scale, we computed the number of circular triads the listener made. A circular triad occurs when, for example, stimulus B is judged to be darker than A, and C is judged to be darker than B but not darker than A. In a subset with 8 stimuli a score of less than 9 out of a maximum of 20 possible circular triads was accepted as a consistent and reliable result (0.05 significance level); in a subset with 9 stimuli this criterion results in less than 14 out of a maximum of 30 circular triads (e.g., Edwards, 1957). Three different explanations of circular triads can be given: (1) the stimuli are almost equal, (2) the semantic scale is not appropriate, which means that the associated acoustical property is not present sufficiently in the stimuli or that the semantic scale is non-linear, or (3) the listener response is not reliable.

Table 12.2. Percentage of accepted judgments on all semantic scales for each subset. The total spectral variance per subset is also given.

subset	musicians	non-musicians	total spectral variance (dB^2)
III	87	83	332
IV	79	64	239
V	90	81	255
VI	90	64	200
VIII	64	54	103
XI	94	92	
average	85	71	

In the case of a subset with many almost equal stimuli a high number of circular triads for all semantic scales has to be expected. In Table 12.2 we give, for all subsets, the percentage of accepted semantic scale results for musicians and non-musicians. The first group was, on the average, more reliable (85 % vs. 71 %). The number of accepted results for each subset shows , especially for the non-musicians, a clear relation with the total spectral variance in the subsets, which is also presented in Table 12.2. This indicates the influence of the degree of dissimilarity between the stimuli on consistency. The song phrase, subset XI, is judged most consistently by all listeners. This is not surprising, since the voice characteristics are much more distinct in a song phrase than in stationary vowels.

Since an inappropriate semantic scale will result in inconsistent results for all listeners, we computed for each semantic scale the percentage of accepted judgments over all subsets. For the musicians the semantic scales with less than 75 % accepted judgments were full-thin, rough-smooth, strong-weak, open-throaty, and dramatic-lyrical. This can be explained for the scales rough-smooth and dramatic-lyrical, since these scales refer to temporal characteristics which are not present in the vowel subsets. For the non-musicians less than 60 % accepted results were obtained for rough-smooth, colorful-colorless, and melodious-unmelodious. Best results were obtained for the semantic scale light-dark for musicians and for the scale clear-dull for non-musicians.

12.3.2 Relations between semantic scales

Since the semantic scales are unlikely to be independent, we investigated their relations in the following way. Per subset, we started with the scores per listener per stimulus per semantic scale. For each listener these eight or nine stimulus scores were rank-correlated for each pair of semantic scales. The resulting correlation matrices of semantic scales were analysed by INDSCAL. In this case the multidimensional scaling technique presents the semantic scales in the object space in which distance is related to dissimilarity: the more distant the positions of two scales, the more dissimilar they are. Differences among listeners in relations between semantic scales are presented in the subject space. To overcome the dependence of the INDSCAL result on the chosen polarity of the semantic scales, each correlation matrix was extended by including all correlations between semantic scales with reversed polarities. After this extension, the versions of a semantic scale with opposite polarities are positioned radially and symmetrically relative to the origin in the object space. In the presentation of the configuration in the object space only that polarity of a semantic scale will be given which is the more easily interpretable.

INDSCAL analyses were performed separately for musicians and non-musicians. We combined the correlation matrices of all listeners for the five vowel subsets. This combination of matrices from different vowel subsets is legitimate because the subject space will reveal whether relations between semantic scales are vowel dependent. This, however, appeared not to be the case for both groups of listeners, which implies that for our stimuli relations between semantic scales have a general validity for a listener, irrespective of vowel and F_0. For the sake of clarity, we subsequently averaged, over the vowel subsets, the position of each listener in the subject space. Results for song phrases were analyzed separately. The results of a two-dimensional INDSCAL analysis for all four cases are presented in Fig. 12.1.

We first consider the subject spaces. We limited ourselves to a two-dimensional solution of INDSCAL, since higher dimensions mostly presented unique relations between semantic scales found for individual listeners. This effect can, for instance, be seen in the second dimension of the non-musician judgments of stationary vowels, which is exclusive to listener 5. The two-dimensional INDSCAL analysis included only a limited part of the total variation in the correlations between semantic scales,

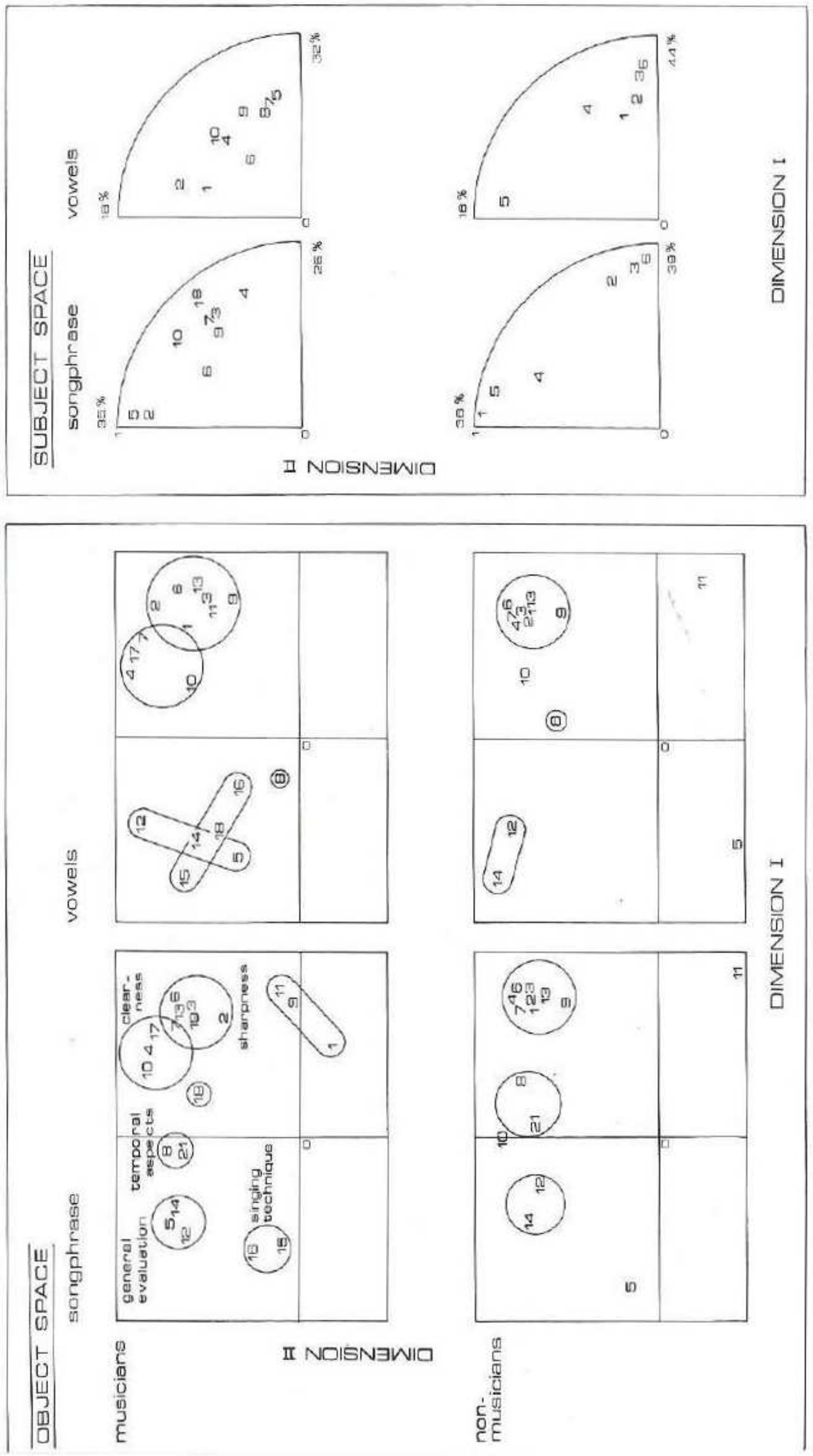

Fig. 12.1. Representation of relations between semantic scales by means of the results of INDSCAL analyses. The I-II plane of both the object space (semantic scales) and the subject space (listeners) is presented for the subset of song phrases and the subsets of stationary vowels, both for musicians and non-musicians. Semantic scale numbers refer to Table 10.1.

varying between 49.1 % (musicians, vowels) and 75.2 % (non-musicians, song phrase). This relatively low percentage explains the deviations of subject weightings from the unit circle. Examination of the subject space shows, especially for non-musicians, a tendency towards a one-dimensional interpretation of the object space, which was either directed towards dimension I or towards dimension II.

Before interpreting the object space, we want to mention some general properties of that presentation: (1) Semantic scales which are close to each other are used more or less synonymously; for easier interpretation clusters of semantic scales are encircled; (2) the distance of a scale to the origin is a measure of its discriminative power (at least in this plane): the closer a semantic scale is located to the origin, the more synonymous it is with its reversed version and the smaller its discriminative power is.

Let us first consider the configuration of semantic scales in the object space for musicians. The clusters of semantic scales are from left to right:

(1) Singing technique: free-pressed (15) and open-throaty (16). For the song phrase the reversed scales 1, 9, and 11: heavy-light, round-angular, and warm-cold can also be considered to belong to this group. Round and warm are probably used as more impressionistic alternatives to the description of a good singing technique. For the stationary vowels, the scales free-pressed and open-throaty cluster with the scales for a general evaluation.

(2) General evaluation: melodious-unmelodious (14), colorful-colorless (12), and full-thin (5). For the song phrase these scales are positively related to scales on singing technique and temporal aspects; for the vowels they overlap the scales on singing technique.

(3) Temporal aspects: vibrato-straight (21) and rough-smooth (8). Rough-smooth is unreliable for stationary vowels (this conclusion was also drawn by Von Bismarck (1974a)). For the song phrase both scales are used independently from scales on singing technique.

(4) Clearness: clear-dull (4), high-low (7), strong-weak (10), and open-covered (17). It is noteworthy that the technical scale open-covered (17) is to a great extent unrelated to other scales on singing technique, namely free-pressed (15) and open-throaty(16).

(5) Sharpness: sharp-dull (3), light-dark (2), shrill-deep (6), and metallic-velvety (13). For the song phrase the scale tenor-bass (19) is also included in this group. The scale soprano-alto was not used in the present analysis, since it was not applied to all vowel subsets. Separate analysis of subsets VI and VIII showed that the scale soprano-alto led to the same judgments as sharp-dull. For stationary vowels the scales light-heavy (1), angular-round (9), and cold-warm (11) can be included in sharpness, too. The scale dramatic-lyrical (18) is used ambiguously. For the song phrase the scale is related to temporal aspects, clearness and sharpness, for the stationary vowels the scale presents a general evaluation.

Most musicians weigh both dimensions of the object space about equally, but for some listeners one of the two dimensions is dominant. In the case of a dominating dimension I, the scales which evaluate temporal aspects are unreliable, while all other scales are used in the same way as sharp-dull, with "sharp" related negatively to general evaluation and singing technique. When dimension II dominates, as was the case for listeners 2 and 5 judging the song phrase, the scales on singing technique are unreliable, while all other scales are judged according to sharpness. For this dimension, sharpness is related positively to general evaluation.

For non-musicians the results for the song phrase and the vowels were very similar and show a further simplification relative to the results of the musicians. The interpretation of the configuration is easy since non-musicians on the whole use either dimension I or dimension II. This implies that, apart from some unreliable scales, most semantic scales are used synonymously, according to sharpness. In view of the position of scales 12 and 14 the subjects only differed in their opinion as to whether sharp had to be associated to unmelodious (when dimension I was used) or melodious (when dimension II was used).

In summary, semantic scales are used very one-dimensionally by most non-musicians and also by some musicians. Most semantic scales are used according to sharpness and the only difference between listeners is whether sharp is positively or negatively associated with a general evaluation of the sound and with singing technique. Most musicians, however, use semantic scales in a more varied way, especially for the song phrase, for which clusters of semantic scales related to singing technique, general evaluation, temporal aspects, clearness, and sharpness can be distinguished. These groups of scales are, however, not independent but can be represented in two dimensions.

12.3.3 Relation between semantic scales and perceptual space

In order to learn how the semantic scales are related to the perception of the stimuli of the different subsets, we used the method of analysis of preference MDPREF (Carroll, 1972). This algorithm computes, from the stimulus scores on the semantic scales, a multidimensional perceptual space in which stimuli are represented as points while in the same space the judgment of a listener on a semantic scale is represented as a vector. The direction of this vector is such that the projection of the stimulus points on the vector correlates maximally (least-squares criterion) with the stimulus scores on the semantic scale concerned. The first dimension of the perceptual space explains most variance in the listeners' judgments, the second most of the remaining variance, etc. The stimulus configuration is normalized so that the variance is equal in all dimensions. The semantic-scale vectors are given unit-length and in graphical presentations of results they are represented by their end points. An example of such a representation is given in Fig. 12.2.

In the non-verbal listening experiment we did not find indications that configurations in the psycho-acoustic perceptual space are different for musicians and non-musicians. Therefore, in the present experiment we combined for each subset the judgments on all semantic scales for both groups. The interpretation of semantic scales, however, may vary from scale to scale and from listener to listener, and will be represented by the semantic-scale vector configuration. Whenever a listener's judgments on a semantic scale included too many circular triads, these data were excluded from

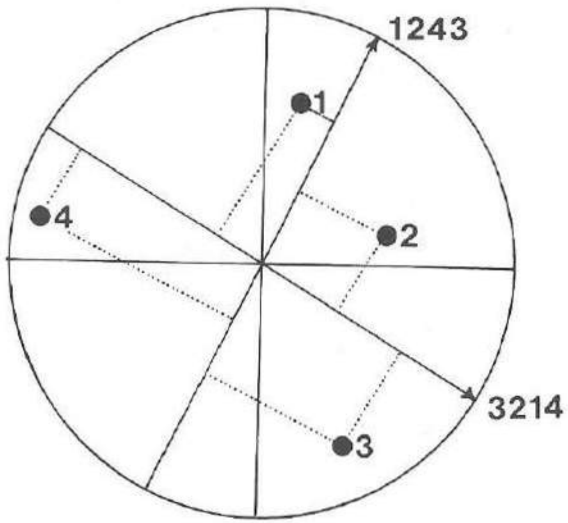

Fig. 12.2. Example of presentation of results of MDPREF analysis. Stimuli are represented as points in a space and semantic scales are represented as vectors on which the projection of the stimulus points gives the best estimate of a listener's judgment.

further computation.

The result of the MDPREF analysis did not allow a generalized description of the interpretation of the semantic scales, due to great inter-individual differences. This will be demonstrated with the help of some examples in Figs. 12.3 and 12.4. In Fig. 12.3a-d typical judgments on the semantic scales are presented for the subset of song phrases. All panels present the same stimulus configuration but different subsets of semantic scale vectors. The first two dimensions of the perceptual space explain 71 % of the total variance in semantic scale judgments. In Fig. 12.3a and Fig. 12.3b all accepted semantic scales for one of the musicians and one of the non-musicians are presented, respectively. The spread of the positions of the semantic-scale vectors for the musician shows that the semantic scales represent various views of the perceptual space. The position of the semantic scales is in good agreement with the general relations between semantic scales for musicians presented in Fig.12.1. The clustering of the semantic-scale vectors for the non-musician demonstrates that most semantic scales, except vibrato-straight (21), are used synonymously, implying a restricted view of the perceptual space.

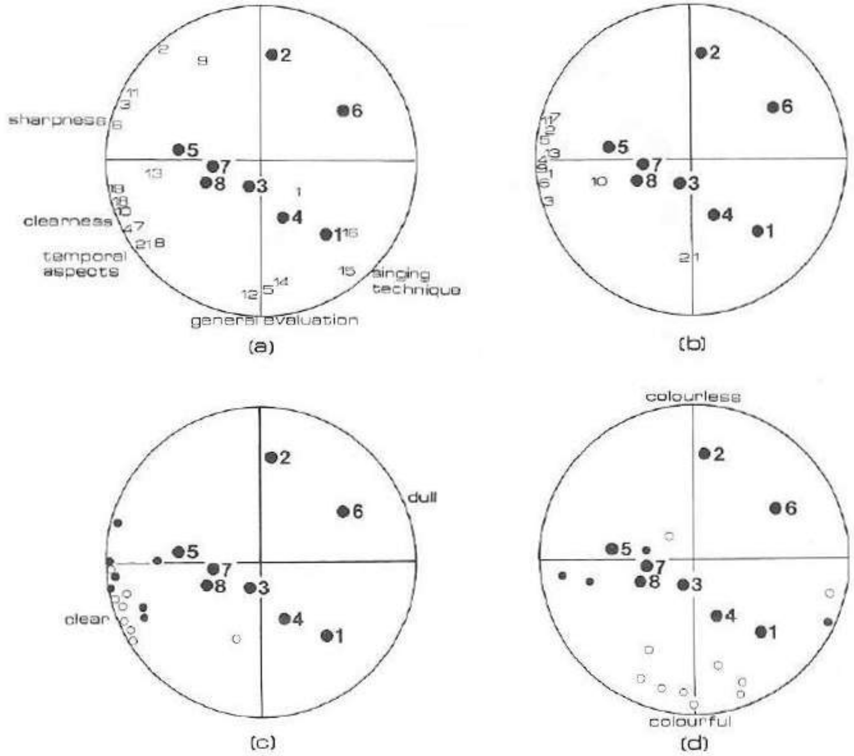

Fig. 12.3. Perceptual space for the song phrases sung by eight male singers (large filled circles, the numbers of which refer to Table 10.1) with various results of semantic-scale vectors:
(a) All accepted semantic scales (numbers refer to Table IV) for one of the musicians.
(b) All accepted semantic scales for one of the non-musicians.
(c) All accepted listeners on the semantic scale clear-dull; small open circles refer to musicians, small closed circles to non-musicians.
(d) All accepted listeners on the semantic scale colorful-colorless; small open circles refer to musicians; small closed circles to non-musicians.

In Fig. 12.3c and Fig. 12.3d, for all musicians (open circles) and non-musicians (filled circles) the directions of the semantic scales clear-dull and colorful-colorless are shown, respectively. The scale clear-dull shows corresponding judgments along the first dimension for all listeners. This is not the case for the scale colorful-colorless, which most musicians judged to be close to the second dimension of the perceptual space, while non-musicians judged again along the first dimension. It can

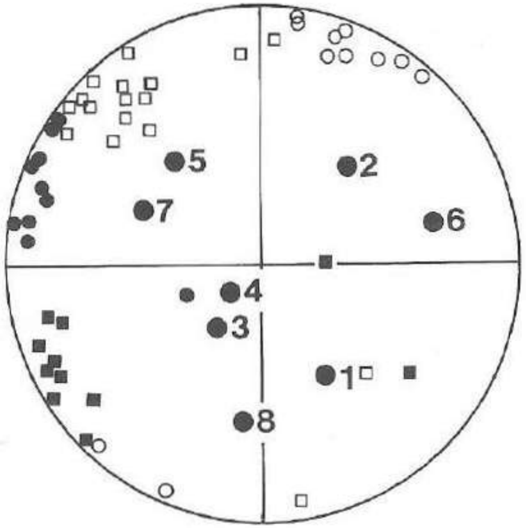

Fig. 12.4. Perceptual space for a subset of vowels /i/, sung by eight male singers at F_o = 220 Hz. All semantic-scale vectors are shown for two musicians (open circles, open squares) and two non-musicians (closed circles, closed squares).

be seen that along dimension I some listeners even have opposite opinions on this scale. The musician and the non-musician for whom the semantic-scale vector is positioned near the origin use this scale in another dimension.

Fig. 12.4 shows the two-dimensional MDPREF solution for subset IV (/i/ sung by 8 male singers at F_o = 220 Hz). The positions of all accepted semantic scales are given for two musicians (open symbols) and two non-musicians (filled symbols). In this example of stationary vowels the clustering of most semantic scales per listener illustrates that the scales are used synonymously but in different ways by the individual listeners. This effect was generally most clearly present in the subsets with stationary vowels.

The application of semantic scales can only be demonstrated in examples such as those given in Figs. 12.3 and 12.4. Due to strongly listener- and vowel-dependent behavior, the results are hard to generalize. Especially when a listener used semantic scales synonymously, it is likely that only one particular perceptual attribute was dominant in the vowel subset, and that all semantic scales were judged according to this attribute. The attribute concerned differed among listeners.

12.3.4 Comparison of vowel configurations derived from non-verbal and verbal listening experiments

The MDPREF analysis of the data of the verbal listening experiment with vowel stimuli resulted in a perceptual vowel configuration for each subset. Since for these subsets vowel configurations were also available from INDSCAL analysis of data from the non-verbal experiments (Chapter 11), we investigated whether these two configurations, derived with quite different techniques, were comparable. For this purpose, we rotated both normalized vowel configurations to maximal congruence. For all subsets the correlation between coordinate values of vowels on matched dimensions was significant beyond the 0.05 level and the coefficient of alination S varied between 0.36 and 0.54. Although the fit between the two configurations was good, there was no one-to-one correspondence between the original dimensions of the two spaces. The significant verbal dimension related to sharpness was, for instance, not immediately represented in the INDSCAL analysis of non-verbal data. This supports the view of non-unique psychological meanings of the dimensions derived by means of INDSCAL in this case. Apart from the difference in the orientation of dimensions, we may conclude that both the non-verbal and the verbal listening experiment resulted in the same configuration of vowels in the psycho-acoustic perceptual space.

12.3.5 Spectral correlates of perceptual dimensions of vowel subsets

The vowel-point configuration derived by means of MDPREF from semantic judgments can be related to the 1/3-oct spectra representation of the vowels. This was done in the same way as in the non-verbal perceptual experiment, using orthogonal rotation to congruence. (We introduced weighting factors for the dimensions of the normalized MDPREF solution which optimized the fit measure S; the direction of the semantic scales was adjusted according to the weighting factors). The dimensionality of the perceptual space was determined by the correlation between each perceptual and corresponding spectral dimension which was significant beyond the 0.05 level. The results of the matching procedure are given in Table 12.3. The dimensionality of each subset and the total amount of spectral variance explained are similar to the results of the non-verbal experiment, with the exception of subset VII (Table 11.1). It can be concluded that our choice of semantic scales was appropriate for describing the psycho-acoustic

Table 12.3. Results of matching vowel configurations in the spectrum space and the optimally weighted verbal perceptual space derived by MDPREF. Ls represents the number of listeners for each subset; S represents the coefficient of alination, D is dimension.

set	F_0 (Hz)	Ls	percentage variance in semantic scale judgments					percentage spectral variance					correlation between perceptual and spectral coordinates along dimensions				S
			D1	D2	D3	D4	Σ	D1	D2	D3	D4	Σ	D1	D2	D3	D4	
III	392	16	58	15	11	-	84	39	28	21	-	88	0.96	0.90	0.91	-	0.48
IV	220	14	54	20	8	4	86	16	27	13	36	92	0.96	0.91	0.95	0.90	0.46
V	220	16	49	17	12	10	88	28	23	22	18	91	0.95	0.90	0.88	0.82	0.51
VI	220	15	57	14	9	-	80	32	31	22	-	85	0.91	0.87	0.95	-	0.53
VII	392	21	33	23	-	-	56	18	14	-	-	32	0.94	0.70	-	-	0.88
XI	s-p	16	57	14	10	5	86										

perceptual space. However, it is not the differences between semantic scales but the interindividual differences in the interpretation of semantic scales which make that conclusion possible.

The good fit between vowel configurations in the verbal perceptual space and the spectrum space allows us to associate properties between the same directions in both spaces and to assign a spectral vector to a perceptual semantic-scale vector. This spectral vector consists of the direction cosines between the original spectral dimensions (related to the sound level in frequency bands) and the vector direction. In Chapter 4 we called the presentation of the direction cosines as a function of the center frequency of the frequency bands the profile of the vector. In the present case, the profile can be considered to represent the spectral variation which underlies perceptual judgments on a semantic scale. Spectral vectors can be derived for all individual semantic scale judgments on stationary vowels. It would be of interest to search for spectral descriptions with a general validity for semantic scales. Unfortunaly, the large interindividual differences in the interpretations of semantic scales, demonstrated previously, do not allow this. However, we can give the corresponding spectral interpretations of the principal dimensions of the verbal perceptual space. These dimensions are determined on the basis of the explained variance in the semantic-scale judgments; the first dimension explains most of this variance, the second dimension most of the remaining variance etc. In

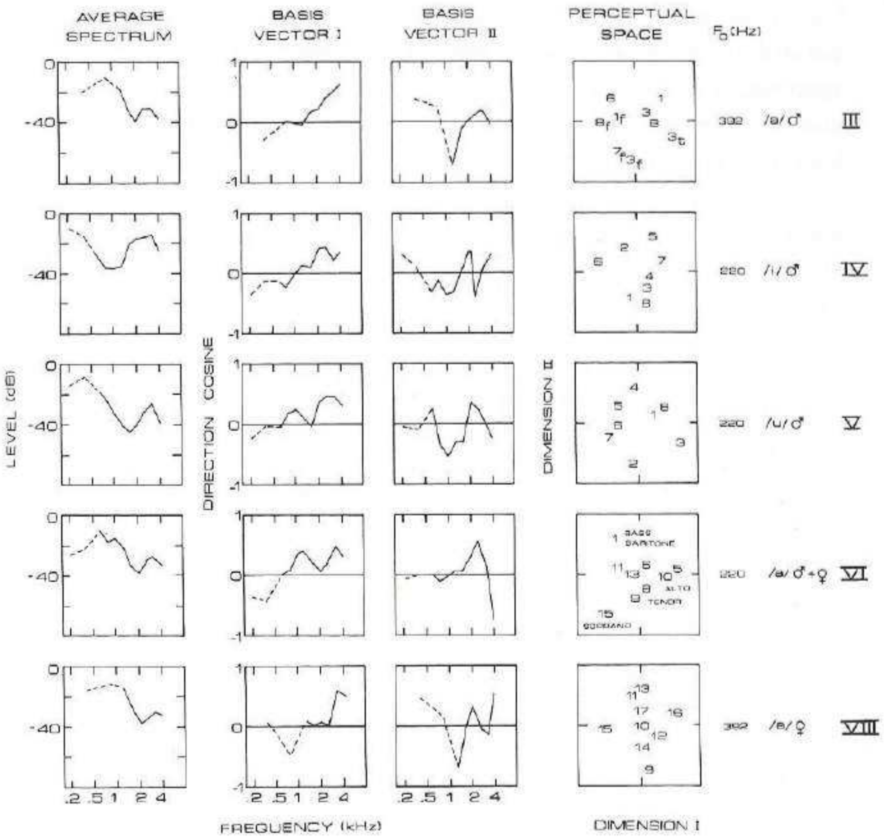

Fig. 12.5. Results of matching vowel configurations in the verbal perceptual space and the spectrum space. Left-hand panels show the grand average spectrum of each subset; the middle panels present the profiles of the spectral basis vectors associated with the first two perceptual dimensions, and the right-hand panels show the vowel configurations in the perceptual space. Numbers refer to singers (see Table 10.1), f indicates falsetto register, t indicates a tenor-like phonation produced by baritone singer 3.

Fig. 12.5 the main results are presented. This figure shows for the five subsets of stationary vowels used in the verbal experiments (1) the average spectrum, (2) the profiles of the first two spectral basis vectors, associated with the matched perceptual dimensions, and (3) the vowel configuration in the plane of the first two dimensions of the perceptual space.

Table 12.3 presents, for all matched dimensions, the percentages of explained variance in semantic-scale judgments and the percentages of explained spectral variance in the vowel stimuli. From this table it can be seen that for most subsets more than half of the total variance in perceptual judgments is covered by the first dimension of the verbal space. Figure 12.5 (panels of the second column) demonstrates that this dimension typically has spectral-slope weighting properties. Spectral slope is independent of the vowel type of a subset and has a general interpretation. This corresponds well with the acoustical properties of the sharpness attribute of timbre, described by Von Bismarck (1974b). It is remarkable, however, that spectral slope also plays the most important role when the corresponding amount of spectral variance is relatively low, as is the case for subset IV (see Table 12.3). For some other subsets, variation in spectral slope coincides with typical properties of the vowel subset: for subset III the configuration in the perceptual space (Fig. 12.5, last column) shows that dimension I contributes highly to the differentiation between falsetto and modal registers (except singer 6, a very "dull" tenor); in subset V the singers 3 and 8 colored the vowel /u/ towards /o/; the spectral effect of this phonemic difference (all formant frequencies of /o/ are higher than those of /u/) is represented along dimension I; in subset VI dimension I of the perceptual space shows that the differentiation between soprano and alto singers has spectral-slope like properties (see also Chapter 4). In Section 12.3.2 it was said that the semantic scale soprano-alto is used in the same way as sharp-dull. This correspondence between soprano and sharp implies that strong higher harmonics in the vowel sounds used in this experiment are associated with soprano singing, which is completely contrary to the actual spectral differences between soprano and alto singers.

The profiles of the second basis vector (see the third column of Fig. 12.5) show that the related perceptual dimensions have no general acoustical interpretation. The properties of these dimensions are probably related to the effects of vowel articulation. For subsets III and VIII, the vowel /a/, the profile of the second basis vector is comparable and weighs

the level of the frequency band of 1.2 kHz. Such a profile differentiates between the phonemes /ɑ/ and /a/; not all singers produced the requested phonemic quality precisely. In subset VI, with combined male and female phonations of /a/, the second dimension weighs the frequency positions of the peaks of higher formants. This property roughly discriminates between tenor and bass singers, as can be seen in the configuration in the perceptual space (Fig. 12.5, last column; see also Chapter 5). For subsets IV and V, the vowels /i/ and /u/ respectively, the second dimension weighs the depth of the spectral valley between lower and higher formants. This dimension is, for the vowel /i/ (subset IV), strongly related to phonemic differences: most vowels /i/ were colored towards /y/, except for the singers 2, 5, and 6. It may be noted that for subsets IV, V, and VIII the perceptual vowel configuration (Fig. 12.5, last column) does not have a relationship to the voice classifications of the singers.

The distribution of the variance in semantic-scale judgments over the perceptual dimensions, and therefore the order of these dimensions, depends on our choice of selected scales. We cannot exclude that a single semantic scale describes a specific perceptual dimension, explaining little variance, while a large number of scales may be focused on one other perceptual dimension, explaining a large amount of variance. In previous sections it has been shown that for the subsets of stationary vowels there is agreement among listeners about scales which describe sharpness, the present first dimension. A detailed study of the data did not reveal another perceptual dimension for which listeners agreed in their description. The present second and higher dimensions therefore mainly rely on the extent to which listeners, unsystematically, use the acoustical properties of these dimensions in their judgments.

In summary, when listeners are requested to judge stationary vowels on a semantic scale, they probably focus primarily on differences in spectral slope between the vowels, even when this difference is smaller than those for the other dimensions (see Table 12.3, subset IV). A large number of different semantic scales, related to sharpness, is judged according to this criterion. For other perceptual dimensions there is no agreement among listeners on verbal attributes.

12.3.6 Spectral correlates of perceptual dimensions of song phrases

The acoustical properties of the song phrase 'Halleluja' are much more complex than the spectral aspects of stationary vowels. Temporal aspects, such as vibrato and vowel duration, may also influence the judgments of listeners. This makes it difficult to relate the perceptual dimensions to all possible acoustical correlates. Since the first /ɑ/ and final /a/ of 'Halleluja' took up more than half of the total phrase duration, we used these two vowels to investigate spectral correlates. The 1/3-oct spectrum of the vowel segments was measured every 10 ms, the resulting 10-ms spectra were normalized for overall sound-pressure level and the average of these spectra was considered to be the representative spectrum of each singer. Subsequently, the resulting configuration of eight points in the spectrum space was matched with the perceptual configuration. Three dimensions showed significant correlations ($p<0.05$). The profiles of the spectral basis vectors associated with the first three perceptual dimensions are shown in Fig. 12.6, together with the grand average spectrum. The first dimension accounts for 45 % of the total spectral variance. The profile of the first basis vector shows that the corresponding perceptual dimension, describing sharpness, is, for song phrases too, associated with spectral-slope-like variation. The profile of the second basis vector strongly weighs the sound level of the frequency bands with center frequencies of 0.8 and 2.5 kHz. This indicates that a positive contribution of this dimension, perceptually associated with full, melodious, and

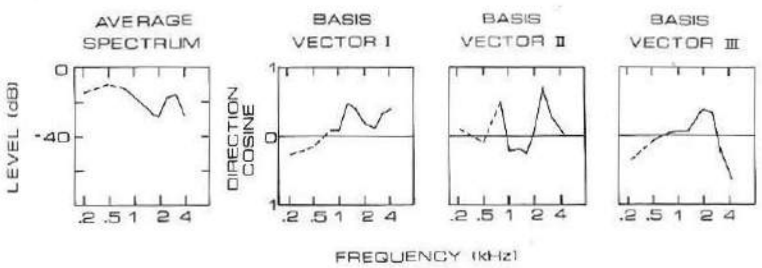

Fig. 12.6. Results of matching the perceptual configuration of song phrases and the configuration of average spectra of the two vowels /ɑ/ and /a/ in 'Halleluja'. The grand average spectrum and the profiles of the first three spectral basis vectors are shown.

colorful, is related to a more open vowel, /a/ instead of /ɑ/, and a higher sound level of the high spectral peak. The latter peak is also known as the singer's formant and described as the origin of the "ring" of the voice (Bartholemew, 1934). The frequency position of this spectral peak is weighed by the third basis vector. No systematically used verbal attribute was associated with this direction.

The amount of spectral variance accounted for by the second and third dimensions was only 9 and 12 %, respectively. Therefore, it may well be possible that other acoustical factors than those present in the average spectrum of the two vowels contribute to these dimensions. Concerning the influence of temporal measures on perceptual judgments, no effect of total phrase duration (tempo) could be established. However, this could be expected since the listeners were requested to ignore this factor. The specifically temporal semantic scale vibrato-straight was judged to a great extent on the basis of the depth of vibrato modulations ($r = 0.74$) and not on vibrato rate ($r = -0.13$). Fig. 12.1 showed that this scale was both positively related to sharpness and to general evaluation (full, melodious, colorful). Therefore, the presence of a good vibrato may contribute as a temporal attribute to these factors.

12.4 Discussion

Up to now no experiments have been reported verifying the equivalence of verbal and non-verbal listening techniques for the measurement of the representation of stationary vowels at the psycho-acoustic level of perception. We confirmed this equivalence up to the limitation of the accuracy of both types of techniques. This can be interpreted as a validation of the measurement technique of pair-wise comparisons and the multidimensional scaling techniques INDSCAL and MDPREF used in the present study.

It has been found from both verbal and non-verbal experiments that a dissimilarity in timbre correlates with a difference in 1/3-oct spectrum for vowels, at least up to a fundamental frequency of 392 Hz. This allows us to investigate properties of vowel representations in the psycho-acoustic perceptual space on the basis of 1/3-oct spectra only, that is, without the need to perform time-consuming perceptual experiments. For vowels sung by professional singers such a study has been reported in Part I.

For all subsets of stationary vowels only sharpness turned out to be a verbal attribute of timbre on which most listeners, regardless of their degree of musical training, agreed in their judgments; they only differed in their evaluation of sharpness: whether sharpness was melodious or not. In conformity with Von Bismarck (1974b), sharpness was found to be acoustically related to the slope of the spectrum. The importance of sharpness in timbre perception was even apparent for subset IV, in which only 16 % of total spectral variance was associated with this factor. These results support von Bismarck's opinion that sharpness may be considered as a fundamental perceptual quality, besides pitch and loudness, of any harmonic complex tone.

As has been said in Chapter 9, for spoken and sung vowels both a psycho-acoustic level and a phonetic level of perception can be distinguished. At the more central, phonetic, level the phonemic identity of a vowel is determined. This level is especially sensitive to formant frequency variation (Klatt, 1982). It may well be possible that a number of verbal attributes of the timbre of vowel sounds refer to this level of perception and describe, for instance, formant frequency deviations from typical average values. The acoustical interpretation of such verbal attributes would then be vowel dependent. Since our subsets each included only one vowel, this kind of timbre description could have emerged from the listening experiments. Although the second perceptual dimension of the vowel subsets, shown in Fig. 12.4, did turn out to be related to vowel-specific acoustical variation, no indications were found that listeners agreed in their verbal description of this variation. This suggests that there are no stable verbal attributes for the phonetic level of perception under the experimental conditions used here.

The present experiments failed to reveal a relationship between the description of timbre of stationary vowels and voice classification (see Fig. 12.5). In all cases both the semantic scales tenor-bass and soprano-alto were used in the same way as sharp-dull: the more high frequency energy, the higher the estimated voice classification. In fact, this first perceptual dimension was unrelated to actual voice classification and even showed a reverse relationship with female voice classification for subset VIII (Fig. 12.5). Whereas such results may be attributed at first sight to the restrictions of stationary vowels which make even musically trained listeners unable to estimate voice classification, the observation persisted to some extent for the phrases sung by male singers. Although in this case the relationship between judgments on the scale tenor-bass (19)

and actual voice classifications was rather good (as an example, see Fig. 12.3), a contradictory result was obtained for tenor singer 6, who had a rather dull voice and was associated with the lowest voice classification. For the song phrases the semantic scale tenor-bass was also highly correlated with the first perceptual dimension and therefore associated with the slope of the spectrum (Fig. 12.6). Listeners seem to relate a shallow spectral slope to tenor voice timbre type and a steep spectral slope to bass voice timbre type. A shallow slope may originate in the spectral effect of high first and second formants (e.g. Fant, 1960) or in a shallow source spectrum. Cleveland (1977) indicated that higher formant frequencies are indeed associated with higher voice classifications in professional male singers. The contradictory result for "dull" tenor singer 6 in the present experiment possibly demonstrated the confusing influence of a steep source spectrum This raises the interesting question of whether perceptual voice classification, based on timbre, has a phonetic basis (formant frequency detection) or a psycho-acoustical basis (sharpness detection). The present results suggest a psycho-acoustic basis which may lead, however, to incorrect judgment of voice classification. Fortunately, many more factors establish voice classification, which obviates a wrong judgment on the basis of timbre alone.

There is, of course, a considerable gap between experiments with electronic sounding stationary vowels and the perception of real singing. The listening experiment with song phrases was a first attempt to bridge this gap. Just as for stationary vowels, sharpness was the most important spectral attribute. The second perceptual dimension (colorfulness) indicated the influence of the relative sound level of the singer's formant. Vibrato was not found to take up a separate perceptual dimension but vibrato quality may enhance judgments on both sharpness and colorfulness, especially when the spectral attributes of these perceptual dimensions are small.

Finally, we should be careful in interpreting semantic scales in terms of acoustical properties in view of the small number of singers in the subsets. Accidental combinations of acoustical characteristics of singers, or their absence, may have influenced the results. Nevertheless, we trust that the main effects, found for most subsets, are likely to have a more general validity.

PART III

13 RETROSPECTION AND PERSPECTIVES

In this chapter we discuss some general aspects of the approach of this study. Firstly, we consider the design of our spectral measurements, the limitations of the interpretation of our spectral analyses, and we suggest the direction into which new experiments may provide more conclusive answers to the questions and hypotheses raised. Secondly, we discuss the implications of the results of our perceptual study on interindividual differences in timbre (Part II) on the inventory of spectral variation measured in professional singing (Part I). Finally, we have to answer the question which may be posed rightly by singers and vocal pedagogues: what can we learn from this study.

13.1 The spectrum of sung vowels

The chosen experimental design of our spectral measurements of sung vowels was based on the observation that in many, often excellent, studies on the singing voice, only a limited number of relevant variables was investigated. When only one singer, a few vowels, a limited range of F_0 values, an unspecified vocal intensity are considered, there always remains some doubt on the generalizability of conclusions at which one arrives. On the other hand, an exemplary setup, limited in variables, but allowing a detailed study, may provide useful ideas on phenomena which are important in singing while the generalizability of results may be investigated in subsequent studies. The present study seems to be unique in the sense that a fairly complete set of the variables vowels, singers (including males, females, registers), modes of singing (including vocal intensity), and fundamental frequency have been chosen for the investigation of their spectral correlates. On the basis of this broad approach we were in a position to discuss the generalizability of more exemplary types of studies and often it turned out that the generalization of conclusions was justified. Perhaps the most surprising deviations were obtained for spectral correlates of female voice classification (Chapter 5) and for the level of the singer's formant (Chapter 8). Spectral correlates of female voice classification were not simply an extension of what was known already for male voice classification, and the sound level of the singer's formant, simply defined as a frequency band level, was not a unique phenomenon in professional male singing, especially in bass singers, but a general voice characteristic which was highly dependent on vowel, voice classification, vocal intensity, and fundamental frequency. There is a need for perceptual studies on the singer's formant to specify its acoustical requirements.

In our study we did not include non-singers and consequently we do not know in how far our results are limited to professional singers. However, a comparison with spectra of spoken vowels, measured by Pols and co-workers, revealed no great discrepancies for configurations in the vowel space (Chapter 4), while differences in the level of the singer's formant could be explained as the effect of different values of overall SPL for sung and spoken vowels (Chapter 8). Nevertheless, more studies, for instance on the spectral differences between singers and non-singers and on the spectral effect of vocal intensity, may provide enough data to judge the generalizability of our results.

The extensiveness of the material necessitated a systematic approach for its analysis. This included averaging procedures in which we chose to focus on the outspoken effects and not to embarrassthe reader with a wealth of details. Another argument for disregarding details was the inherently rather crude character of 1/3-octave spectra. Nevertheless, the main results are interesting. We established the contributions of the factors "vowels," "singers," and "modes of singing" and their interactions to total spectral varation in singing, as a function of F_0: spectral differences between singers and modes of singing did not depend very much on F_0 and preponderated over differences between vowels for F_0 values higher than about 659 Hz (E5) (Chapter 3). We showed the shrinking of the vowel space with increasing F_0, in which the difference between front and back vowels was preserved the longest (Chapter 4). We also showed spectral correlates of voice classification and demonstrated that female singers differed from each other in another way than males, which has a possible explanation in dominant glottal differences between female singers and dominant morphological differences between male singers (Chapter 5). We showed the dominating influence of spectral-slope-like variation in modes of singing (Chapter 6) and we presented a detailed analysis of the level of the singer's formant (Chapter 8).

It should be stressed that our efforts to relate the results mentioned to voice production are of a speculative nature because 1/3-oct spectra do not have a direct interpretation in terms of voice production. Still, we trust that the indications presented in this study are clear enough to justify a future attempt for its verification. This may be illustrated by our results on voice classification. We found that the differences between alto and soprano singers were very sensitive to the modes of singing soft and loud (Fig. 5.4). Since in soft and loud singing a glottal component is certainly involved, this indicates that such a factor may play a role in the explanation of spectral differences between female singers. The same type of reasoning applies to the spectral difference between male modal and falsetto registers, an accepted laryngeal phenomenon, the spectral effect of which resembles the spectral differences between alto and soprano voices (Fig. 5.2). Spectral differences between male singers, allowing an interpretation in terms of morphological differences of the vocal tract in conformity with Cleveland (1977), were of a different nature than those among female singers (Figs. 5.1 and 5.2), which indicates that the morphological differences which differentiate between male singers do not simply extend to female singers. These three indications, probably too weak when considered one by one, seem strong enough together to hypothesize a glottal component in the difference between alto and soprano singers.

This example is typical since it shows on the one hand the weakness of 1/3-oct spectral analysis in its failure to present conclusive results and interpretations in terms of voice production; on the other hand it also shows that from a combination of data, obtained under various conditions, new and, hopefully, fruitful ideas may arise. Whereas 1/3-oct spectral analysis needs many data for its interpretation, it is the power of this approach that these data can be measured easily, fast, and fully automatically. It is not accidental that the broad experimental design used in this

the classical approach on the basis of the acoustical theory of speech production: phonations at moderate and high F_0 values, strong source-spectrum variation, and acoustic interaction between glottis and vocal tract would provide great analytical problems. Therefore, a cooperation between both approaches may prove to be most successful: Outlines presented by 1/3-oct spectrum analysis, such as presented in this study, followed by detailed experiments and analysis on the basis of models of voice production.

If our hypotheses on the origin of spectral differences are correct, this implies that glottal factors have a great impact on differences between singers and modes of singing and may answer a lot of questions on the representation of interindividual and intra-individual differences in singing and possibly also in speech. Yet this contradicts Sundberg's (1973) remark that "the development of voice timbre in voice training would be a matter of learning a special articulation rather than having the vocal cords vibrate in a very special way". However, on the basis of investigations on the vocal sound source, Gauffin and Sundberg (1980) introduced more recently the difference between "flow" and "pressed" phonations. Flow and pressed phonations probably are no discrete adjustments of the vocal folds but the extremes of a continuous range of possible adjustments. Pressed phonation is certainly at the bad end of this range, but the other end may contain a large range of acceptable vocal-fold adjustments useful for the control of (variation in) voice timbre. We speculated that a certain quality of flow phonation may be a prerequisite for the perceptual existence of the singer's formant.

13.2 The timbre of sung vowels

Our perceptual experiments allowed the conclusion that the results of a spectral analysis, based on 1/3-oct spectra, can be interpreted from a perceptual point of view (Chapter 11). On this basis we may combine the conclusions from our perceptual experiments using semantic bipolar scales with the results of the spectral analysis of vowels sung by professional singers. These conclusions are (1) that sharpness is the only attribute of the timbre of steady-state sung vowels upon which musicians and non-musicians agree and which is roughly related to the slope of the spectrum, and (2) that for second and higher perceptual dimensions, which are related to vowel-dependent spectral properties of a vowel, no common description exists. Spectral-slope-like variation, from a characteristically low amplitude of the fundamental and large amplitude of high harmonics to large amplitude of the fundamental and small amplitudes of the high harmonics, forms a dominant part of the spectral effects of vocal effort (loud-soft), of register (modal-falsetto), and of female voice classification (alto-soprano). These effects are probably all related to the glottal level and verbally described by sharp (steep glottal pulse) versus dull (shallow glottal pulse) or synonymous terms.

The spectral analysis of modes of singing allowed an interpretation of a second spectral dimension in terms of dark versus pressed singing as intended by the singers (Chapter 6). This dimension was vowel dependent and probably should be associated with pharyngeal volume and larynx height. However, in the verbal perceptual experiments (Chapter 12) there was no agreement on a second perceptual dimension among listeners. It may be possible that this was the effect of the design of the perception experiments which only included the factor singers but not modes of singing. New experiments may study whether the intention of a singer to produce a timbre type such as dark or pressed is reacted to by listeners in the same terms.

Listeners could not correctly judge voice classification on the basis of the spectrum of stationary vowels, nor on the basis of a song phrase. The higher voice classification was always associated with a shallower spectral slope, which is incorrect for alto-soprano differences and may interfere with spectral effects of a glottal origin so that a "dull" tenor can be associated with a bass timbre type. Considering the practical importance of voice classification in singing, a detailed study of these contradictory phenomena seems very desirable.

13.3 Some implications for singing practice

Finally, I address myself to singers and singing teachers. They will probably look upon this book in embarrassment, unable to understand much of it. At first sight many topics may sound of interest to them but a second look will tell them that answers to the questions raised are not easy to trace and to understand for those with little affinity with and knowledge of phonetic science. Therefore, let us summarize some results of this study which may be of direct importance to them.

(1) Our results suggest that the end of the confusion in the terminology on timbre is still far away in singing; probably that era will never come. Even professional listeners, who use the terminology more reliably than non-musicians, may completely disagree over the acoustical interpretation of terms, or may describe the same acoustical phenomena in quite different terms. On the average, the terms clustered into the categories: singing technique (such as open-throaty), general evaluation (such as melodious-unmelodious), temporal factors (such as straight-vibrato), clearness (such as clear-dull), and sharpness (such as sharp-dull). These categories were not independent. For instance, melodious was related to both a good vibrato and a good singing technique. A fairly common opinion among listeners about terms was only found for the timbre difference sharp-dull, which was associated with the slope of the spectrum (amplitudes of the higher partials relative to the amplitude of the fundamental). Although the lack of agreement would make one feel sadly, the observation that for most aspects of timbre a unified, objective terminology simply does not exist should lead us to conclude that we have to be tolerant in discussions. Our verbal incapacity may have been recognized by those singing teachers who simply demonstrate and ask the pupil to imitate the example. On the other hand it is possible that for a group of singers, who belong, for instance, to the same school, a terminology may be functional inside, but probably not always outside this group.

(2) Our inability to give unequivocal verbal judgments seems to extend to voice timbre types. The more important the higher partials, the greater the tendency of professional listeners to assign a higher classification to a voice. This is incorrect for female voices because in alto singers the higher partials are far more prominent than in soprano singers, and is also incorrect when high partials come, or, in contrast, do not come into prominence due to glottal factors. Therefore, voices should never be classified

on the basis of timbre alone, but on a wide variety of acoustical and physiological measurements.

(3) We found many indications that glottal factors should not be underestimated in singing. They play a role in register differences, variation of vocal effort, and in timbre differences between female singers. Modal register, greater vocal effort, and alto voices therefore exhibit the stronger high partials, resulting in a "richer" voice quality.

(4) As a general rule, male singers seem to differ among themselves mainly in the shape and length of the vocal tract, especially the pharynx; bass singers have a large pharyngeal volume and tenor singers have a small pharyngeal volume. Female singers seem to differ mainly in the way the vocal folds vibrate: A faster glottal closure for alto singers and a slower glottal closure for soprano singers. The average difference between male and female singers is based on the shape and length of the vocal tract.

(5) It may be possible to vary the speed of glottal closure in the falsetto register, which could be the basis of a soprano-like sound for untrained falsetto (slow glottal closure or even incomplete closure), and of an alto-like sound for a counter-tenor (fast glottal closure).

(6) When pitch was raised, spectral (and timbre) differences between vowels decreased, which is the basis of the decreased intelligibility of vowels at high pitches. The difference between the vowel group /a/, /ɑ/, /ɔ/, /u/, and the group /i/, /y/, /e/, /ɛ/, /œ/ were the longest to remain. Spectral differences between singers and between modes of singing were much less affected by increasing pitch and preponderated in female singing at fundamental frequencies higher than 659 Hz (E5).

(7) Differences in modes of singing are, in the first place, probably highly related to glottal factors (vocal-fold adduction) and are verbally described by sharp-dull and loud-soft. In the second place, differences between modes of singing are related to larynx height and pharyngeal volume, which are verbally described by dark-pressed. Besides this morphological aspect, pressed singing is often also associated with strong vocal fold adduction.

(8) The variation in timbre an individual singer may realize by different modes of singing is comparable to the variation in timbre between different voices. This implies that a similar timbre can be realized by voices of quite different classifications. Hence, a baritone may be able to sing with tenor quality and vice versa, although it remains questionable whether this is always good for the voice.

(9) The sound level of the singer's formant (the "ring" of the voice) is influenced by many factors. A high sound level is found in most voices, with the possible exception of soprano voices, at high vocal effort and low fundamental frequencies. Without disputing Sundberg's articulatory explanation of the singer's formant as the acoustical effect of a lowered larynx, glottal factors also seem to contribute a great deal to the prominence of the "ring". It cannot be excluded that some quality in vocal fold vibration is a prerequisite for the perceptual presence of the singer's formant.

SUMMARY

In this study the variation in the spectrum and in the timbre of sung vowels was studied. In the acoustical investigations (Part I) the intervocalic (different vowel) and intravocalic spectral variation (different singer or mode of singing), measured in 1/3-octave frequency bands, was analyzed. In the perceptual investigations (Part II) the timbre dissimilarities between sung vowels, were mapped in a multidimensional space, using data from listening experiments. Both approaches are closely related. We demonstrated that at least up to a fundamental frequency (F_0) of 392 Hz, the small dissimilarity in timbre of the same vowel sung by different singers could be predicted on the basis of differences in the 1/3-octave spectra (Chapter 11). The results of the analysis of the 1/3-oct spectrum measurements, which are much easier to obtain experimentally than are perceptual data, could be interpreted, therefore, from a perceptual point of view.

For the acoustical study, Part I, we recorded 3888 sung vowels, which are highly representative of all possible timbres of sung vowels. Fourteen professional singers (2 basses, 2 baritones, 3 tenors, 3 altos, 2 mezzo-sopranos, 2 sopranos) sang nine different vowels, each in nine different modes of singing (neutral, light, dark, pressed, free, loud, soft, straight, extra vibrato) at six F_0 values, ranging from 98 to 880 Hz.

For a general impression of the extent to which the main effects of the factors "vowels," "singers," and "modes of singing", and their interactions contributed to spectral variation, we determined in Chapter 3 for each F_0 value the contribution of main effects and interactions to total spectral variance. The spectral variance due to the factor "vowels" dominated up to F_0 = 659 Hz; spectral differences between singers and modes of singing preponderated for higher F_0 values.

An explanation of the multidimensional representation of 1/3-octave spectra is presented in Chapter 4. A relationship is shown to exist between the results of an analysis of vowel spectra on the basis of the acoustical theory of speech production, using formant frequencies, and on the basis of 1/3-octave spectra. A first order correspondence is demonstrated. On this basis the properties of configurations of vowel points are studied up to F_0 = 880 Hz.

Characteristic spectral differences between singers are dealt with in Chapter 5. It was concluded that spectral differences among male singers are different from those among females, which probably points to the spectral effect of differences in shape and length of the vocal tract for males, and of differences in glottal pulses for females. In this chapter spectral differences between the modal register and the falsetto register are also presented.

Different modes of singing are, to a great extent, related to spectral slope (Chapter 6). Extremes are the modes of singing "soft" (pianissimo) and "loud" (fortissimo), which have their major origin at the glottal level. Less prominent are spectral differences related to larynx height and pharyngeal volume, which are present in the modes of singing "dark" and "pressed".

In Chapter 7 it is shown that the four modes of singing "soft," "light," "neutral," and "loud," together constitute the major cause of variation in overall sound-pressure level (SPL). A second cause of this variation is to be found in the average differences in overall SPL between singers. An analysis of spectral differences associated with variation in overall SPL is presented; the spectral effects were vowel dependent.

An analysis of the sound level of the so-called singer's formant in Chapter 8 results in the conclusion that large variation (up to more than 30 dB) may occur due to differences between singers, vowels, modes of singing, and F_0. It is striking that the sound level of the singer's formant relative to overall SPL is kept highly constant in each mode of singing, by tuning F_0 and regulating overall SPL (vocal effort). For the same F_0 value the level of the singer's formant was, on the average, considerably lower only with soprano singers, while singers with other voice classiffications had roughly the same singer's formant level.

Part II deals with the timbre of the same vowel, sung by different singers, studied in two different types of listening experiments. Stimuli were based on recordings of advanced student singers. In the first experiment dissimilarity in timbre was investigated in a non-verbal way (Chapter 11); in the second experiment differences in timbre were judged verbally on 21 semantic bipolar scales (Chapter 12).

Both experiments resulted in comparable representations of timbre in a multidimensional space, which excellently matched the representation in a spectrum space based on the 1/3-octave spectra of the stimuli. Additionally, the verbal experiment made it possible to investigate the terminology

of timbre. Highest agreement among listeners was found for the semantic scale "sharp-dull," which was related to the slope of the spectrum. Whereas non-musicians did not differentiate this semantic scale from the other scales, the reverse was true for the musicians, especially when a song phrase was judged. For the musicians, the semantic scales clustered into the following categories: singing technique, general evaluation, temporal aspects, clearness, and sharpness; these categories were not independent.

SAMENVATTING

In deze studie is de variatie in het spektrum en in het timbre van gezongen klinkers behandeld. In het akoestische onderzoek (deel I) werd de binnen en tussen klinkers optredende spectrale variatie (gemeten in 1/3-octaaf frekwentiebanden) geanalyseerd. In het perceptieve onderzoek (deel II) werd met behulp van luisterproeven het timbre van gezongen klinkers in kaart gebracht. Beide aanpakken zijn nauw verwant: we toonden aan dat minstens tot een grondfrekwentie van 392 Hz de ongelijkheid in het timbre van eenzelfde klinker, gezongen door verschillende zangers, zeer goed voorspeld kan worden op basis van het verschil in 1/3-octaaf spektrum (Hoofdstuk 11). De resultaten van de analyses van de 1/3-oct spectra kunnen daarom geïnterpreteerd worden vanuit het oogpunt van de waarneming van timbre. Het voordeel is dat 1/3-oct spectra in experimenteel opzicht veel eenvoudiger te verkrijgen zijn dan gegevens over timbreperceptie.

Voor de akoestische analyses (deel I) maakten we opnamen van 3888 gezongen klinkers die representatief moeten worden geacht voor een zeer groot aantal mogelijke timbres in zang. Daartoe zongen veertien professionele zangers (2 bassen, 2 baritons, 3 tenoren, 3 alten, 2 mezzo-sopranen, 2 sopranen) negen verschillende klinkers, elk met negen verschillende manieren van zingen (neutraal, licht, donker, geknepen, vrij, luid, zacht, strak, extra vibrato) op grondfrekwenties (F_0) van 98 tot 880 Hz.

Om een algemene indruk te krijgen van de bijdrage van de hoofdeffekten van de faktoren "klinkers", "zangers", en "manieren van zingen", en hun interakties, wordt in Hoofdstuk 3 voor elke grondfrekwentie hun bijdrage tot de totale spektrale variantie bepaald. De aan de faktor "klinkers" gerelateerde spektrale variantie blijkt te domineren tot F_0 = 659 Hz; voor hogere F_0 waarden domineren spektrale verschillen tussen zangers en manieren van zingen.

Nadat de meerdimensionale representatie van 1/3-octaaf spektra is uitgelegd wordt in Hoofdstuk 4 een relatie gelegd tussen de resultaten, in de vorm van formantfrekwenties, uit een analyse van klinkerspektra op basis van de akoestische theorie van spraakproduktie en de resultaten van een analyse op basis van 1/3-octaaf spektra. Een eerste-orde overeenkomst wordt aangetoond. Op basis hiervan worden eigenschappen van klinkerkonfiguraties

tot F_0 = 880 Hz bestudeerd.

Karakteristieke verschillen in spektrum voor verschillende zangers worden in Hoofdstuk 5 onder de loep genomen. Dit leidt tot de conclusie dat spektrale verschillen tussen mannelijke zangstemmen belangrijk anders zijn dan spektrale verschillen tussen vrouwelijke zangstemmen, wat waarschijnlijk wijst op verschillen in de vorm en lengte van het aanzetstuk bij mannen en op verschillen in stembandpuls bij vrouwen. In dit hoofdstuk komen tevens spektrale verschillen tussen het modale register en het falsetregister aan de orde. Verschillende manieren van zingen blijken grotendeels gerelateerd te zijn aan de helling van het spektrum (Hoofdstuk 6). De steilste helling heeft het spectrum van zachte (pianissimo) zang, de vlakste die van luide (fortissimo) zang, welke in belangrijke mate op stemplooinivo hun oorsprong hebben. Minder geprononceerd zijn de spektrale verschillen die met de hoogte van het strottenhoofd en het keelvolume samenhangen: donkere en geknepen zang.

Hoofdstuk 7 laat zien dat vier manieren van zingen (zacht, licht, neutraal en luid) de belangrijkste bron vormen van verschillen in het totale geluidsdruknivo in gezongen klinkers. Op de tweede plaats komen verschillen tussen zangers. Een analyse van spektrale verschillen die met variatie in het totale geluidsdruknivo samenhangen wordt gegeven, waarbij klinkerafhankelijke aspekten aan de orde komen.

Een analyse van het geluidsdruknivo van de zogenaamde zangersformant leidt in Hoofstuk 8 tot de konklusie dat hierin grote variatie kan optreden door verschillen tussen zangers, klinkers, manieren van zingen en F_0. Opvallend was dat het relatieve nivo van de zangersformant ten opzichte van het totale geluidsdruknivo voor elke manier van zingen zeer constant gehouden werd door een afstemming van de geluidsintensiteit en F_0. Het gemiddelde geluidsdruknivo van de zangersformant was alleen voor sopranen beduidend lager dan voor alle andere stemklassifikaties.

Deel II behandelt de resultaten van een tweetal typen luisterexperimenten naar timbreverschillen van eenzelfde klinker die door verschillende zangers werd gezongen. Hierbij werd gebruik gemaakt van opnamen van gevorderde conservatoriumstudenten. Zowel musici als niet-musici participeerden als luisteraars. In het eerste type luisterexperiment werd de ongelijkheid in timbre op een niet-verbale wijze onderzocht (Hoofdstuk 11); in het tweede type experiment werden timbre-verschillen verbaal beoordeeld op 21 semantische bipolaire schalen (Hoofdstuk 12). In het verbale experiment werden niet alleen klinkers maar

ook liedfrasen beoordeeld om het perceptieve effekt van temporele variatie op timbre te onderzoeken.

De beide experimenten leverden voor de klinkers een vergelijkbare weergave van timbre in een meer-dimensionale ruimte op. Het verbale experiment maakte het daarnaast mogelijk om relaties tussen beschrijvingswijzen van timbre, zoals helder, dof, kleurrijk, te onderzoeken. De grootste overeenstemming tussen luisteraars bestond over de semantische schaal scherp-dof, die globaal gerelateerd bleek te zijn aan de helling van het spektrum. Waar niet-muzikaal geschoolde luisteraars verder weinig differentiatie tussen semantische schalen aanbrachten, was dit voor musici wel het geval, vooral bij de beoordeling van een liedfrase. De semantische schalen groepeerden zich rond: zangtechniek, algemene evaluatie, temporele aspekten, helderheid en scherpte; deze groepen waren echter niet onafhankelijk.

APPENDICES

APPENDIX A

A DETAILED DESCRIPTION OF INTRA-VOCALIC SPECTRAL VARIATION

Intra-vocalic spectral variation can be divided into spectral variation due to differences between singers and spectral variation due to differences between modes of singing. For the material described in Part I, spectral characteristics of intra-vocalic variation are given in Figs. A1-A9 separately for all nine vowels. The spectral characteristics of each vowel are described for each F_0 value and for male and female singers separately.

The panels of each figure are organized as follows:
(1) The panels of the left half show results for the male singers, the panels of the right half those for the female singers, (2) each row of panels present results for the same F_0 value, from 98 Hz (upper row) to 880 Hz (lower row), and (3) for each F_0 value, separately for males and females, a row of four panels is presented showing the average spectrum and three basis vectors associated with (a) spectral differences between singers, (b) spectral variation as the effect of differences in overall SPL, and (c) spectral variation due to the modes of singing dark and pressed, respectively. In addition, for male singers at F_0=392 Hz the basis vector which describes the main spectral difference between the modal and the falsetto register is presented too.

The following computational procedures have been used:
AVERAGE SPECTRUM. For each vowel and each F_0 value the average spectrum of the total number of spectra (=number of singers x nine modes of singing) is shown. The effect of the basis vectors presented can be estimated with reference to this average spectrum.
BASS-TENOR/ALTO-SOPRANO. These panels show the basis vector per F_0 value which is associated with to the main spectral differences between singers

for the vowel concerned. A positive coordinate value along the associated direction indicates spectra of bass or alto singers; a negative coordinate value those of tenor or soprano singers. The computation is essentially equal to the one used in Chapter 5 for the description of the spectral differences between singers (averaged over vowels). In the present case, the singer-related dimension was derived separately for each vowel in such a way that the projection of singer points on the dimension was similar for all nine vowels (using an iterative procedure of orthogonal rotation to congruence). The spectral variation due to differences between singers was mainly two-dimensional. The dimension presented explained on the average between 48 % (F_0 = 392 Hz, females) and 68 % (F_0 = 98 Hz) of the variance related to differences between singers; the first two dimensions between them explained between 72 % (F_0 = 220 Hz, males) and 87 % (F_0 = 98 Hz).

MODAL-FALSETTO. The second dimension for each vowel, which described most of the variance remaining after the first dimension between average singer spectra, showed for F_0 = 392 Hz (males) the difference between phonations in the modal and the falsetto register (compare Fig. 5.2). The basis vector of this dimension is presented too.

After the spectra had been normalized for the average spectrum of each singer, the remaining spectral differences were associated with modes of singing for each vowel. We distinguished spectral differences which are related to variation in overall SPL and remaining differences which on the whole described the effect of the modes of singing dark and pressed.

LOUD-SOFT. Using the procedure outlined in Chapter 7, we computed the basis vector of the direction in the spectrum space which was related to the spectral effect of variation in overall SPL. We chose this basis vector instead of the one derived by principal-components analysis (see Chapter 6) because of its formal definition which is less dependent on statistical properties of the data. The direction associated with the basis vector presented explained between 41% (F_0 = 98 Hz) and 86% (F_0 = 880 Hz) of total spectral variance due to modes of singing. A positive coordinate value along this direction indicates spectra associated with the higher overall SPL.

DARK-PRESSED. The dimension which described most of the remaining spectral variance showed the spectral differences between the dark and pressed modes of singing (compare Figs. 6.3-6.5). A positive coordinate value along this second dimension is associated with the darker phonations. The combination of the two spectral dimensions of modes of singing explained between 66%

(F_0 = 98 Hz) and 93% (F_0 = 880 Hz) of total spectral variance due to modes of singing.

To facilitate the study of the rather complex figures we advise the reader to focus attention on the vowels /a/ and /i/, which are highly representative of back and front vowels, respectively.

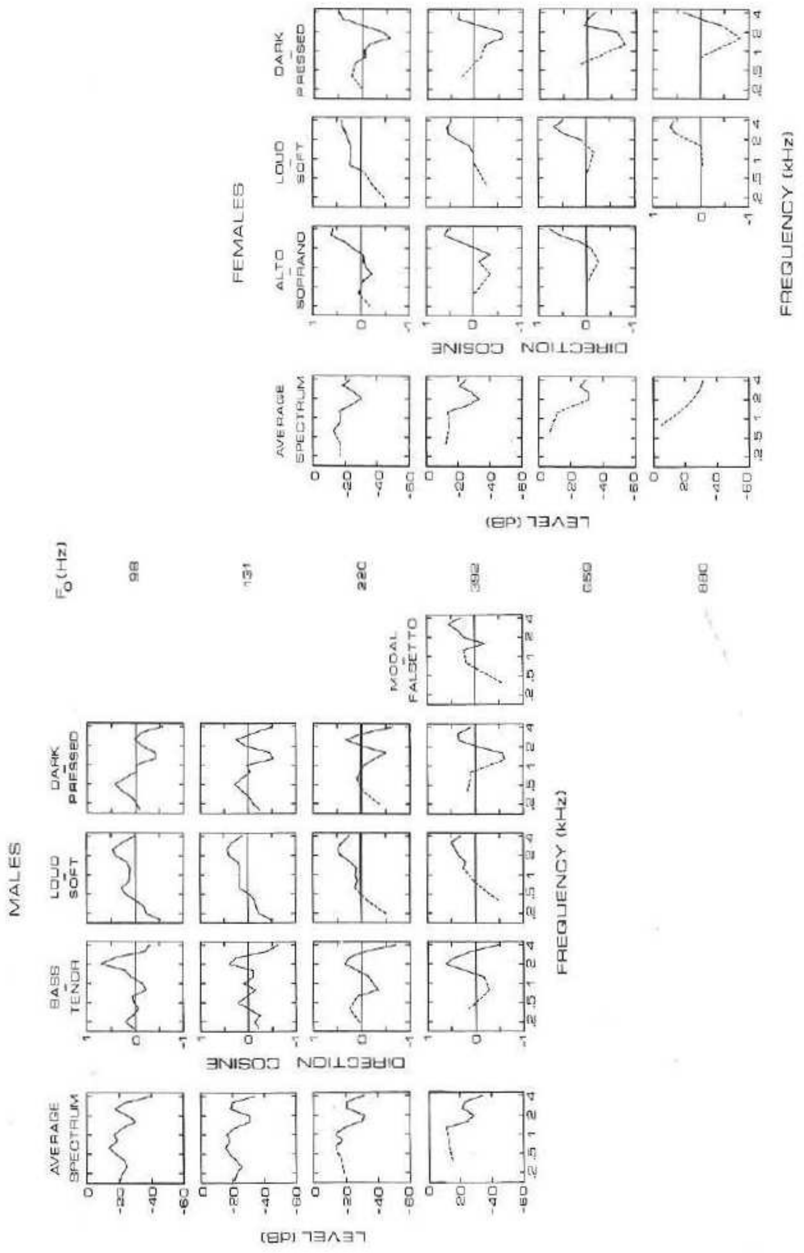

Fig. A1. Characteristics of spectral variation of the vowel /a/. For an explanation see the text.

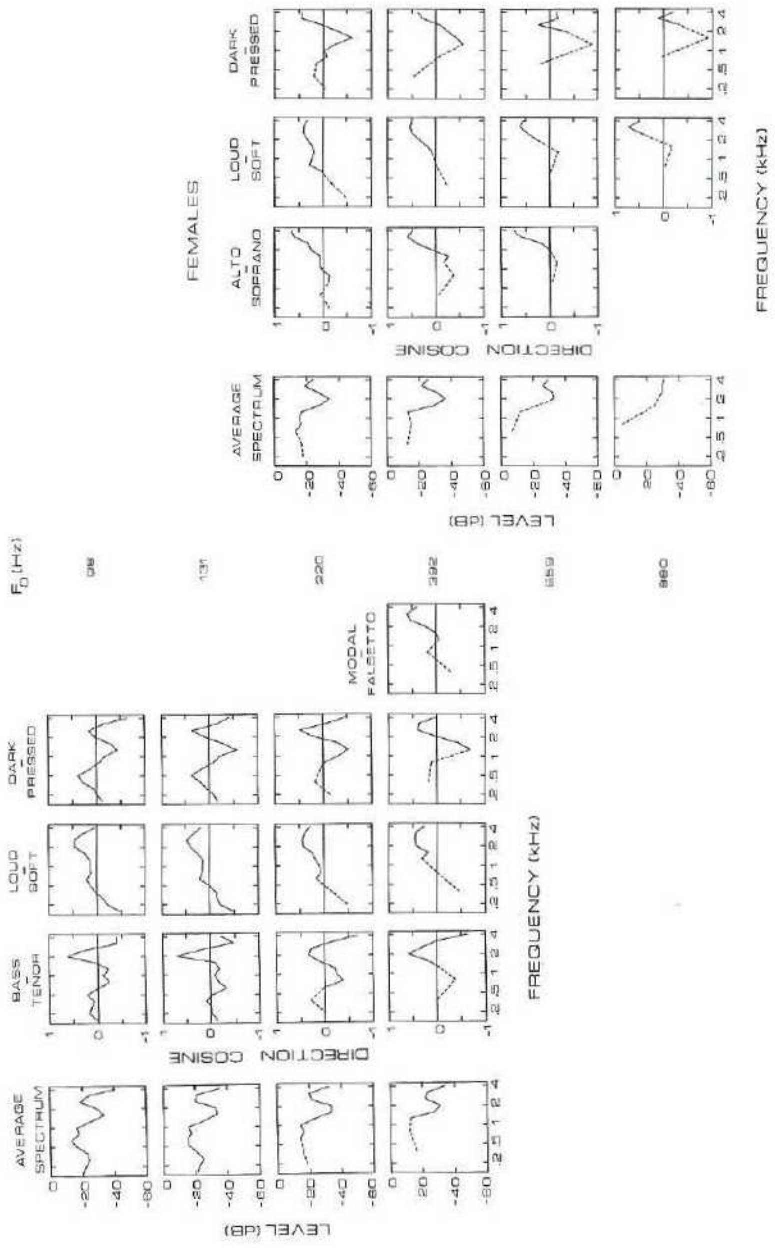

Fig. A2. Characteristics of spectral variation of the vowel /ɑ/. For an explanation see the text.

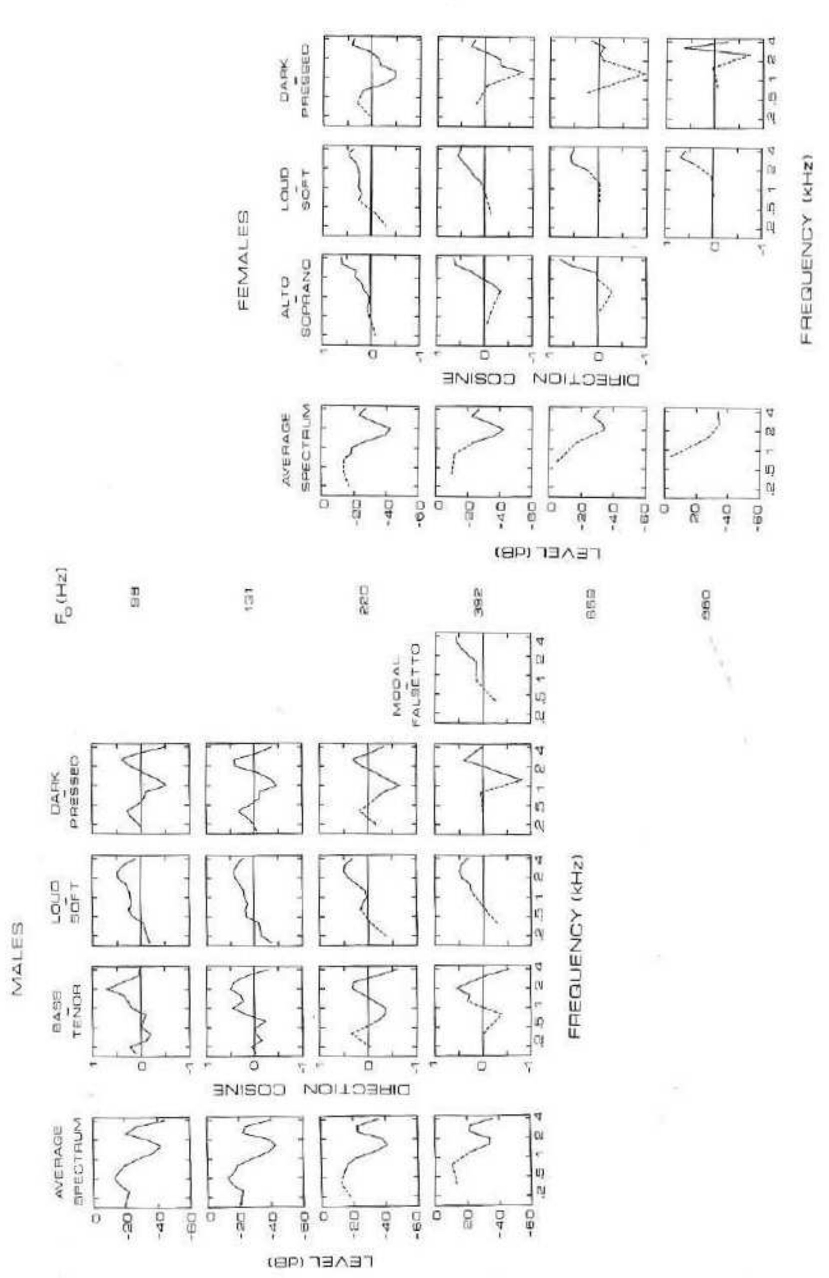

Fig. A3. Characteristics of spectral variation of the vowel /ɔ/. For an explanation see the text.

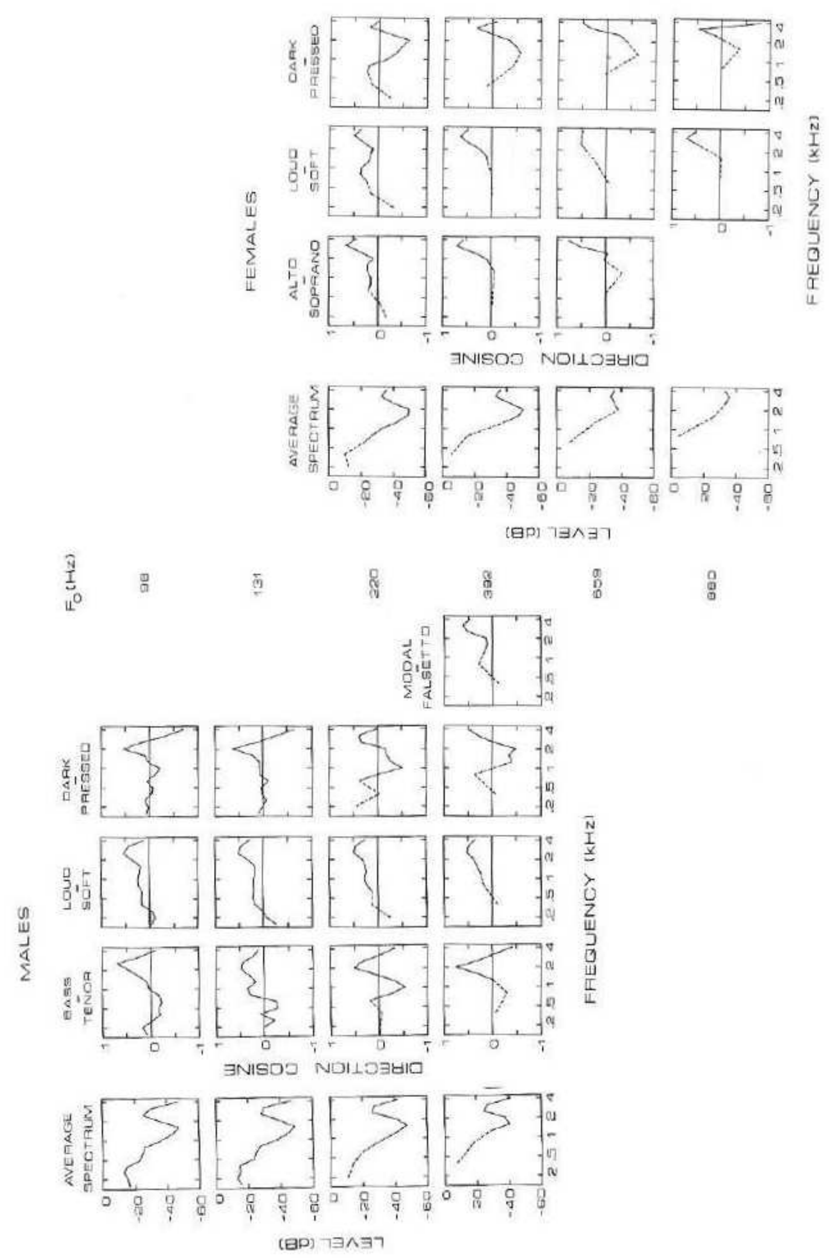

Fig. A4. Characteristics of spectral variation of the vowel /u/. For an explanation see the text.

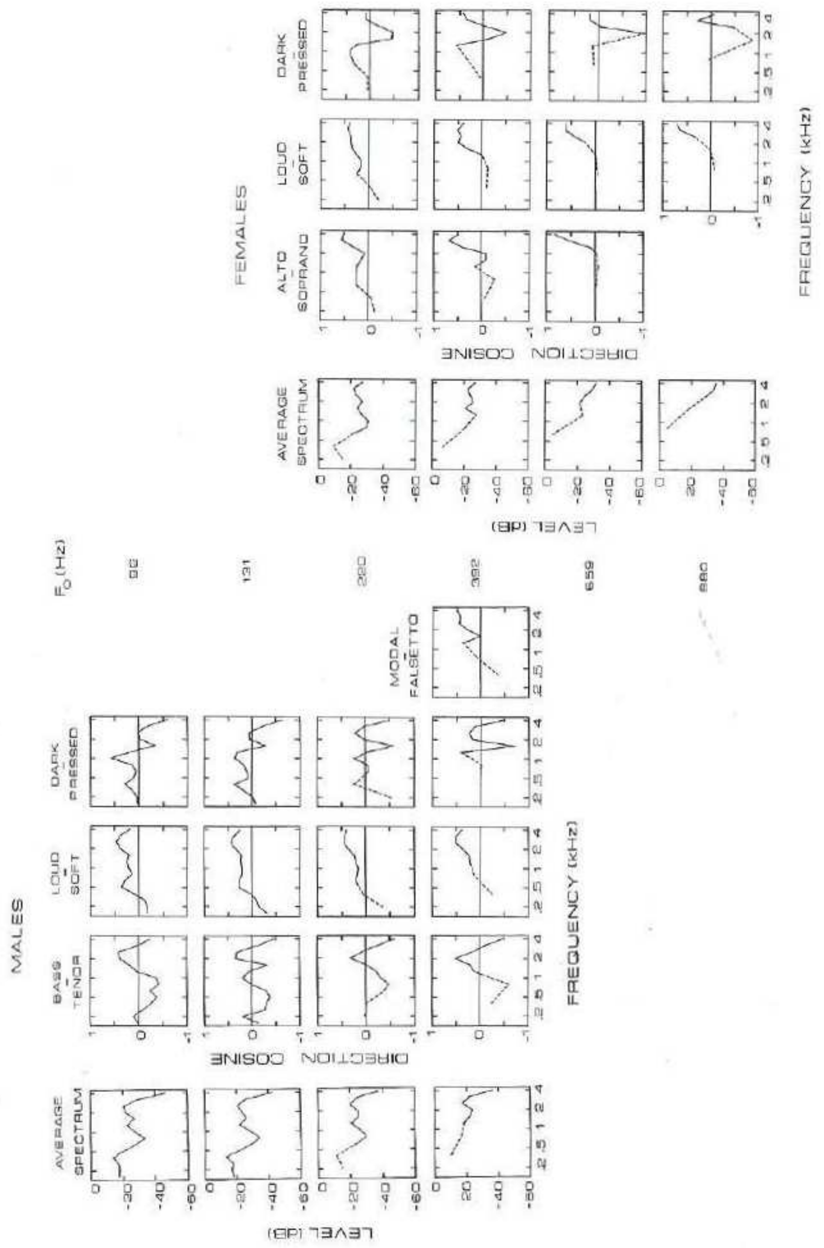

Fig. A5. Characteristics of spectral variation of the vowel /œ/. For an explanation see the text.

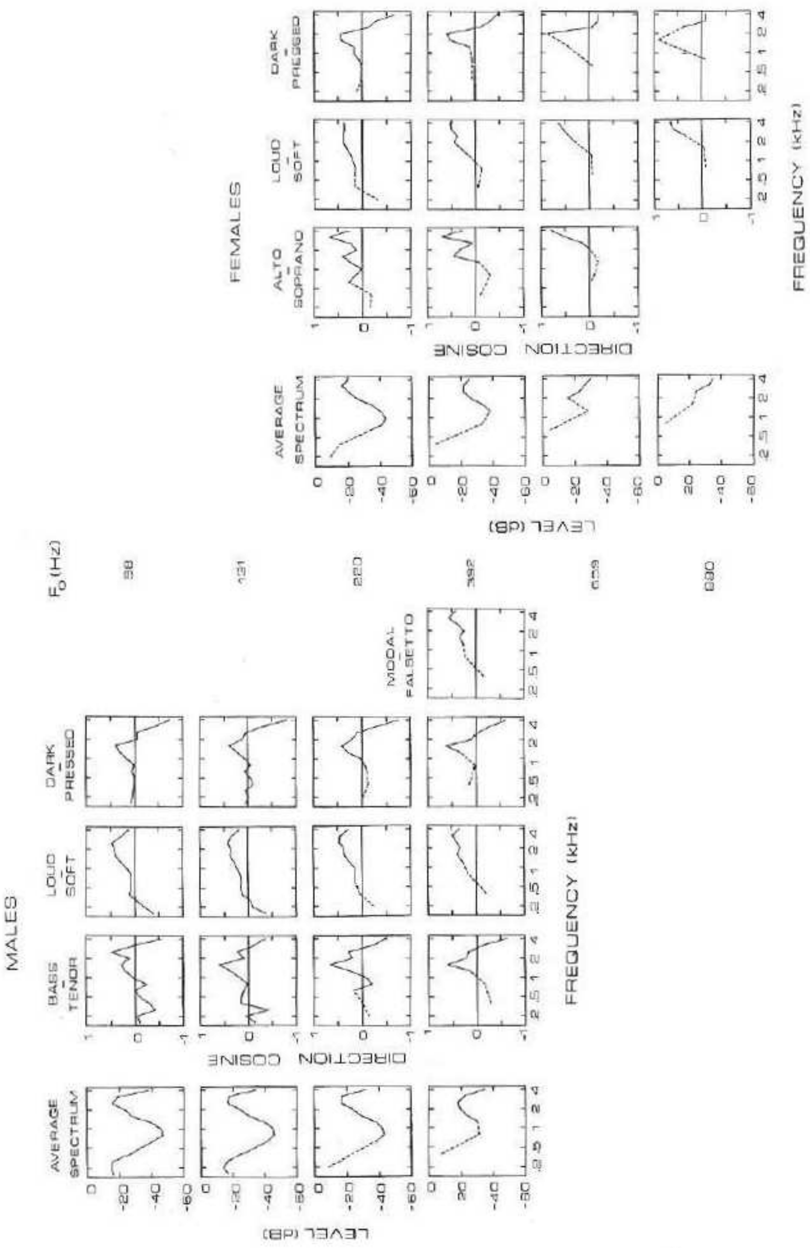

Fig. A6. Characteristics of spectral variation of the vowel /i/. For an explanation see the text.

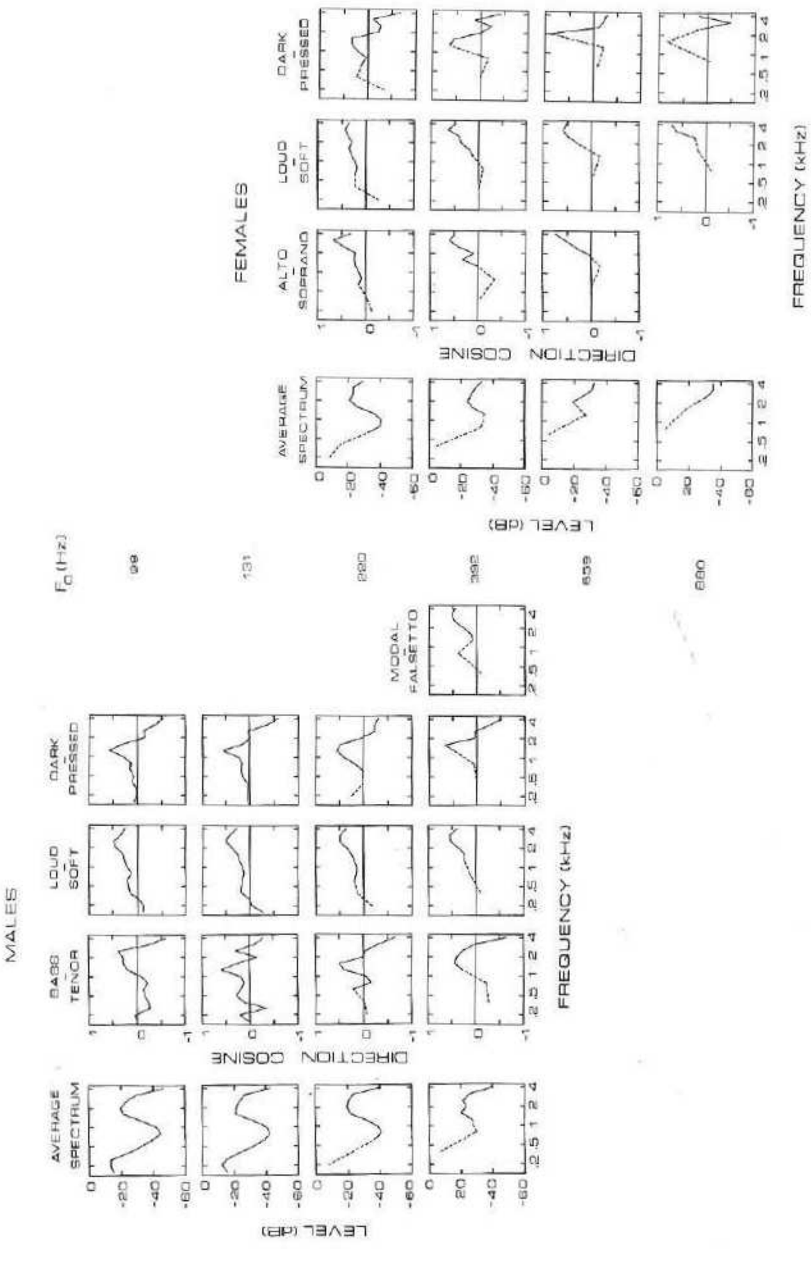

Fig. A7. Characteristics of spectral variation of the vowel /y/. For an explanation see the text.

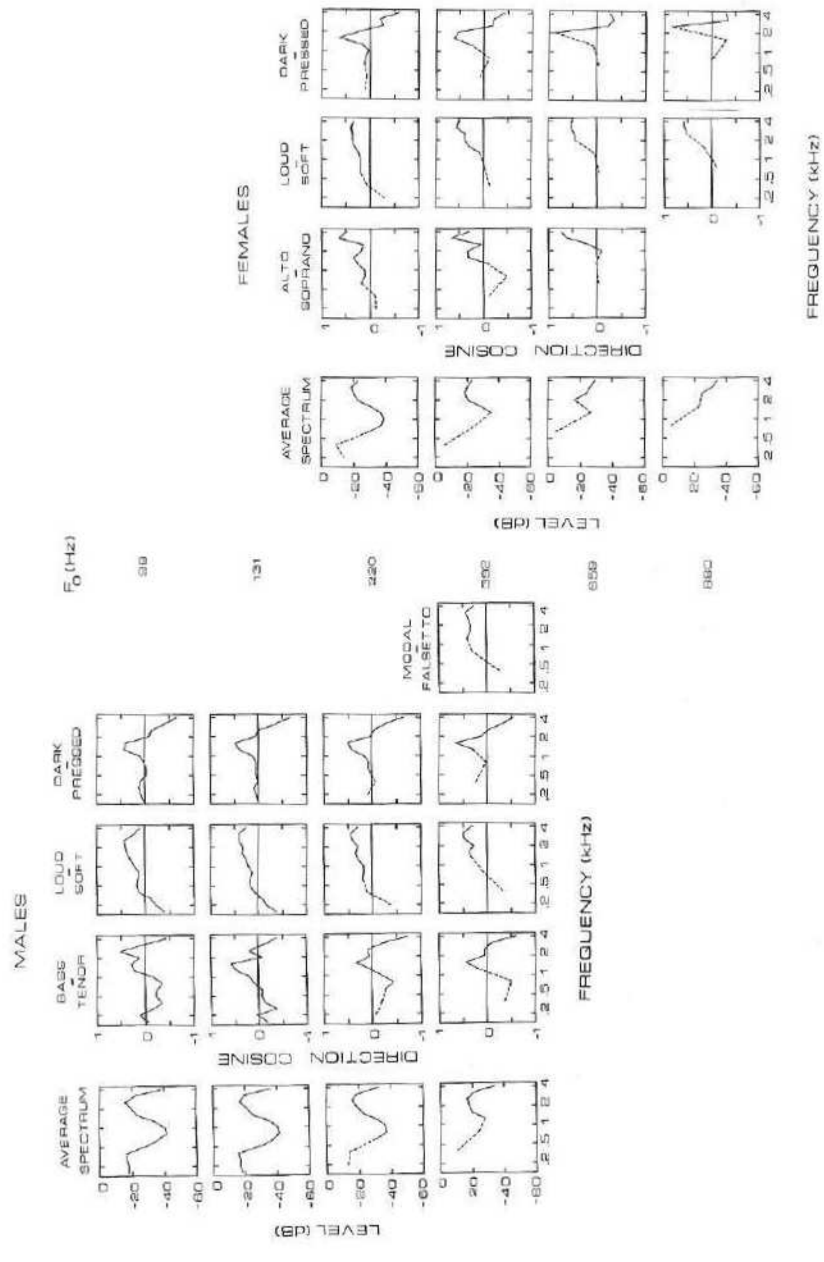

Fig. A8. Characteristics of spectral variation of the vowel /e/. For an explanation see the text.

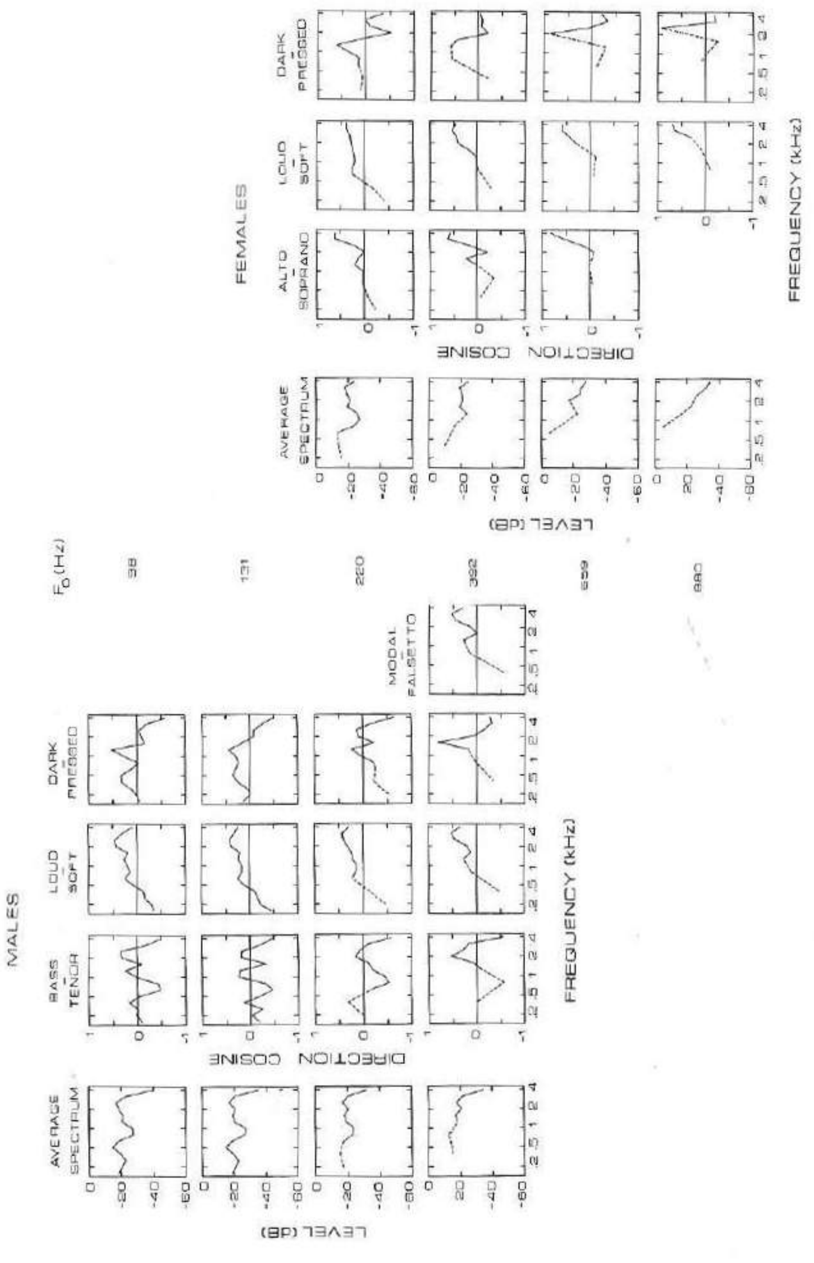

Fig. A9. Characteristics of spectral variation of the vowel /ɛ/. For an explanation see the text.

APPENDIX B

REPRESENTATION OF MODES OF SINGING IN THE FORMANT SPACE AND IN THE SPECTRUM SPACE

In this appendix the relationship between spectrum space and formant space with respect to the representation of the effect of modes of singing is elaborated. The configuration of nine vowels, sung in nine modes of singing by a bass singer at F_0 = 98 Hz, is presented in Fig. B1 in a three-dimensional (sub)space of both spaces. The vowels, of which the average data (averaged over modes of singing) have already been analyzed in Chapter 4, are represented in the spectrum subspace which allowed the best match between both vowel configurations. For the sake of clarity, the vowel points are not marked separately but as the contour of the points of modes of singing for each vowel. Correlation coefficients were 0.89, 0.89, and 0.76 for dimension I, II, and III, respectively. The percentage of explained variance in the spectrum space dimensions was 43.1, 31.4, and 9.9 %, respectively.

Both configurations show that, although the average vowel positions are very similar (see also Fig. 4.1), there are differences in the contours related to the modes of singing: they are wider in the spectrum space. This implies that more spectral variation within vowels is represented in this spectrum subspace than can be attributed to variation in F_1, F_2, and F_3, and that an effect of other factors must be assumed. A detailed study of the configurations revealed that most spectral variation related to variation in overall SPL was uniquely represented in the I-II planes of both spaces. In the formant space front vowels tend to vary more in F_1 due to different modes of singing, while back vowels show a covariation of F_1 and F_2. This is predicted by an articulatory model (Lindblom and Sundberg, 1971) for the spectral effect of a wider jaw opening, combined with a change from rounded to spread lip shape, articulatory movements which are to be expected with increasing vocal effort. In the I-II plane of the spectrum space, the contours are wider and more parallel than in the formant space, and they probably demonstrate the extra effect of the presence of a glottal component in variation of vocal effort. That this component is primarily present in the I-II plane makes a distinction in the spectrum space

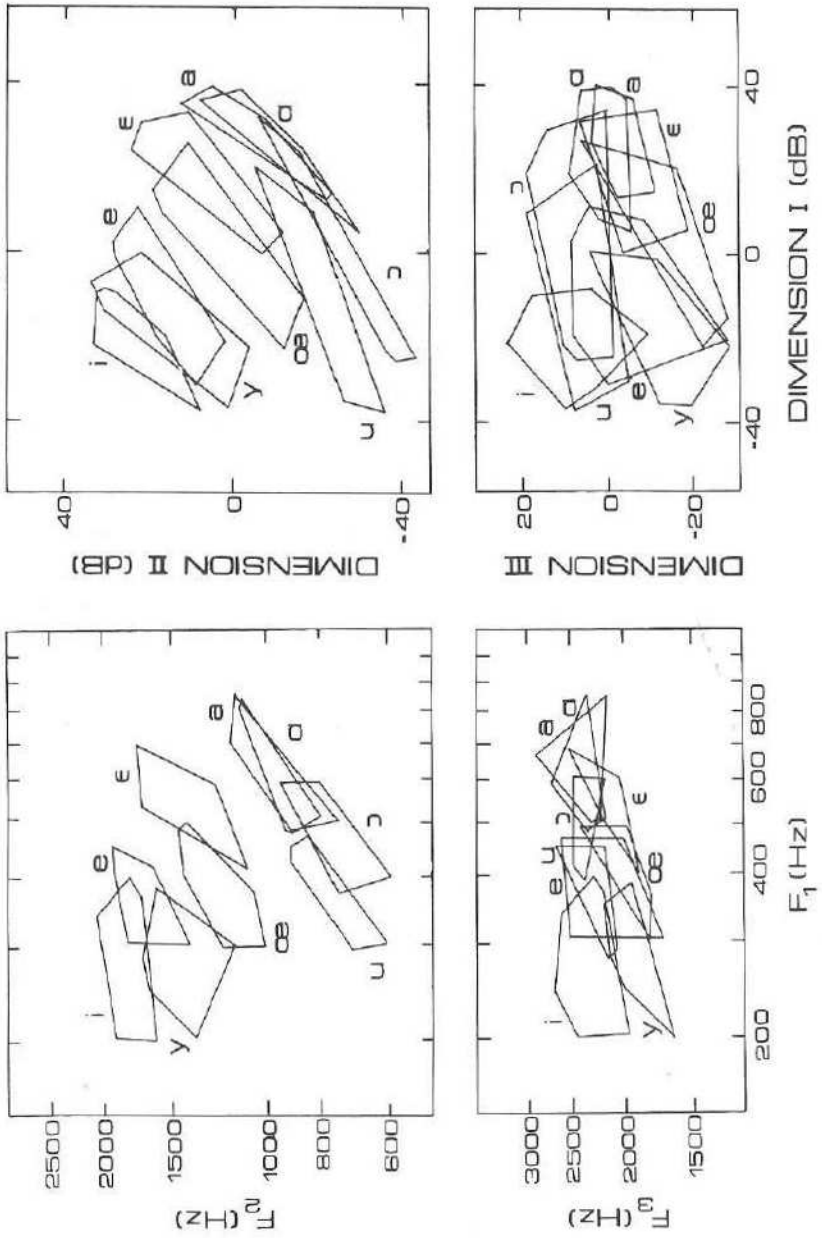

Fig. B1. Representation of vowels, sung by a bass singer at F_o = 98 Hz, in the formant space (left panels) and in the spectrum space (right panels). The contours include data points of the nine modes of singing.

between articulatory and glottal origins of spectral variation impossible, although the glottal component may prevail in intra-vocalic variation.

The variation in the third dimension (Fig. B1) was also much more pronounced in the spectrum space than in the formant space. This may be the spectral effect of a covariation of F_3, F_4, and F_5. If this is the case, the spectral variation associated with F_3 will be partly combined with the spectral variation associated with the higher formants in the third dimension of the spectrum subspace. This dimension will then show more variation than can be expected for F_3 alone. The presence of the singer's formant probably provides these conditions.

In summary, a three-dimensional subspace of the spectrum space exists which explains most of the variation in average vowel positions as well as variation due to modes of singing and different singers (not demonstrated). The three dimensions of this subspace are, in a first-order approximation, related to logarithmic F_1, F_2, and a combination of F_3 upto F_5 axes, respectively. A glottal component in spectral variation due to mode of singing shows up in the same subspace of the spectrum space, introducing a mix-up with spectral variation of articulatory origin.

APPENDIX C

DATA ON OVERALL SPL

Average values of overall SPL (in dB) for modes of singing, singers, and vowels as a function of fundamental frequency. Microphone distance 0.3 m.

	Males						Females			
F_0 (Hz)	98	131	220	392 modal	392 falsetto		220	392	659	880
Modes										
Soft	72	74	77	87	81		72	82	94	90
Light	78	80	84	94	84		78	87	97	100
Dark	84	85	90	101	92		84	92	102	112
Neutral	87	90	91	101	87		84	92	101	105
Free	88	91	92	103	95		84	91	102	108
Straight	85	86	89	102	90		83	89	98	103
Vibrato	88	91	93	103	96		86	84	104	106
Pressed	85	86	90	101	92		86	93	101	109
Loud	90	96	100	108	101		93	99	108	118
Singers										
1 bass	85	88	94	104		8 alto	85	93	102	
2 bass	88	90	94	98	92	9 alto	83	93	105	
3 baritone	83	85	90	104	89	10 alto	83	88	98	
4 baritone	81	89	92			11 m-soprano	75	83	94	
5 tenor		86	88	97		12 m-soprano	89	96	103	104
6 tenor		81	85	98		13 soprano	83	93	101	
7 c-tenor			87		92	14 soprano	86	91	103	107
Vowels										
/i/	83	85	88	100	93		81	92	101	105
/y/	82	85	88	100	94		81	93	102	105
/u/	83	85	89	99	93		83	92	101	105
/e/	84	87	90	100	90		84	90	102	105
/œ/	84	87	91	99	89		84	90	101	106
/ɔ/	85	88	91	100	90		85	90	100	105
/ɛ/	85	87	91	101	89		84	90	100	105
/ɑ/	86	88	91	101	91		84	91	100	106
/a/	86	88	91	100	89		84	90	100	105
Grand-average	84.3	86.6	89.8	100.2	90.9		83.4	90.9	100.8	105.8

REFERENCES

Ågren, K., and Sundberg, J. (1978). "An acoustical comparison of alto and tenor voices," J. of Research in Singing 1(3), 26-32.

ASA (1960). "American Standard Acoustical Terminology", New York, Definition 12.9, Timbre, 45.

Bartholemew, W.T. (1934). "A physical description of 'good voice quality' in male voice," J.Acoust.Soc.Am. 6, 25-33.

Benedini, K. (1980). "Klangfarbenunterschiede zwischen tiefpass gefilterten harmonischen Klängen," Acustica 44, 129-134.

Bladon, R.A.W., and Lindblom, B. (1981). "Modelling the judgment of vowel quality differences," J.Acoust.Soc.Am. 69, 1414-1422.

Bloothooft, G., and Plomp, R. (1984). "Spectral analysis of sung vowels. I. Variation due to differences between vowel, singers, and modes of singing," J.Acoust.Soc.Am. 75, 1259-1264.

Boves, L. (1984). "The phonetic basis of perceptual ratings of running speech," Doct. Dissertation (Catholic University, Nijmegen).

Carlson, R., and Granström, B. (1979). "Model predictions of vowel dissimilarity," Speech Transmission Laboratory, Quarterly Progress and Status Report 3-4, 84-104.

Carroll, J.D. (1972). "Individual differences and multi-dimensional scaling," in Multi-dimensional scaling I, edited by R.W. Shepard, A.K. Romney, and S.B. Nerlove (Seminar, New York).

Carroll, J.D., and Chang, J.J. (1970). "Analysis of individual differences in multi-dimensional scaling via an N-way generalization of 'Eckart-Young' decomposition," Psychometrica 35, 283-319.

Chew, S.L., and Gottfried, T.L. (1981). "Identification of steady-state sung vowels," J. Acoust. Soc. Am. 69, S94.

Cleveland, T.F. (1977). "Acoustic properties of voice timbre types and their influence on voice classification," J.Acoust.Soc.Am. 61, 1622-1629.

Colton, R.H. (1972). "Spectral characteristics of the modal and falsetto registers," Fol. Phoniatr. 24, 337-344.

Colton, R.H. (1973). "Vocal intensity in the modal and falsetto register," Fol. Phoniatr. 25, 62-70.

De Bruyn, A. (1978). "Timbre-classification of complex tones," Acustica, 40, 108-114.

Dmitriev, L., and Kiselev, A. (1979). "Relationships between the formant structure of different types of singing voices and the dimensions of supraglottic cavities," Fol. Phoniatr. 31, 238-241.

Donovan, R. (1970). "The relationship between physical analysis of sounds and the auditory impression of their vowel and tone quality in tenor singing," Acustica 23, 269-276.

Edwards, A.L. (1957). "Techniques of attitude scale construction," (Appleton, New York).

Fagel, W.P.F., van Herpt, L.W.A., and Boves, L. (1983). "Analysis of the perceptual qualities of Dutch speaker's voice and pronunciation," Speech Communication 2, 315-326.

Fant, G. (1960). "Acoustic theory of speech production," Mouton, The Hague.

Fant, G. (1975). "Non-uniform vowel normalization," Speech Transmission Laboratory, Quarterly Progress and Status Report 2-3, 1-19.

Fant, G. (1981). "The source filter concept in voice production," Speech Transmission Laboratory, Quarterly Progress and Status Report 1, 21-37.

Flach, M. (1964). "Uber die Unterschiedliche Grösse der Morgagnischen Ventrikel bei Sängern," Folia Phoniatrica 16, 67-74.

Gauffin, J., and Sundberg, J. (1980). "Data on the glottal voice source behavior in vowel production," Speech Transmission Laboratory, Quarterly Progress and Status Report 2-3, 61-70.

Gottfried, T.L. (1980). "Identification of sung vowels," J. Acoust. Soc. Am. 68, S100.

Harman, H.H. (1967). "Modern factor analysis," (The University of Chigago, London).

Hollien, H. (1983). "The puzzle of the singer's formant," in Vocal fold physiology edited by D.M. Bless and J.H. Abbs (College-Hill, San Diego).

Hussler, F., and Rodd-Marling, Y. (1976). "Singing: the physical nature of the vocal organ," (Faber and Faber, London).

Isshiki, N., Ohamura, H, Tanabe, M, and Morimoto,M. (1969). "Differential diagnosis of hoarseness," Folia Phoniatrica 21, 9-19.

Kakusho, O., Hirato, H., Kato, K., and Kobayashi, T. (1971). "Some experiments of vowel perception by harmonic synthesizer," Acustica 24, 179-190.

Karlsson, I.A. (1984). Proc. of the 10th International Congress of Phonetic Sciences, edited by M.P.R. van den Broecke and A. Cohen (Foris, Dordrecht), p.167.

Karnickaya, E.G., Mushnikov, V.N., Slepokurova, N.A., and Zhukov, S.J. (1975). "Auditory processing of steady-state vowels," in Auditory analysis and perception of speech, edited by G. Fant and M.A.A. Tatham (Academic Press, London).

Klatt, D.H. (1979). "Perceptual comparisons among a set of vowels similar to /ae/: Some differences between psychophysical distance and phonetic distance," J. Acoust. Soc. Am. 66, S86(A).

Klatt, D.H. (1982). "Prediction of perceived phonetic distance from critical-band spectra: a first step," Proc. ICASSP, Paris, 1278-1281.

Klein, W., Plomp, R., and Pols, L.C.W. (1970). "Vowel spectra, vowel spaces, and vowel identification," J. Acoust. Soc. Am. 48, 999-1009.

Lindblom, B., and Sundberg, J. (1971). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. 50, 1166-1179.

Lingoes, J.C., and Schönemann, P.H. (1974). "Alternative measures of fit for the Schönemann-Carroll matrix fitting algorithm," Psychometrica 39, 423-427.

Nierop, D.J.P.J. van, Pols, L.C.W., and Plomp, R. (1973). "Frequency analysis of Dutch vowels from 25 female speakers," Acustica 29, 110-118.

Nord, L. and Sventelius, E. (1979). "Analysis and prediction of difference limen data for formant frequencies," Speech Transmission Laboratory, Quarterly Progress and Status Report, 3-4, 60-72.

Plomp, R. (1970). "Timbre as a multidimensional attribute of complex tones," in Frequency analysis and periodicity detection in hearing, edited by R. Plomp and G.F. Smoorenburg (Sythoff, Leiden).

Plomp, R. (1975). "Auditory analysis and timbre perception," in Auditory analysis and perception of speech, edited by G. Fant and M.A.A. Tatham (Academic, London).

Plomp, R., Pols, L.C.W., and Geer, J.P. van de (1967). "Dimensional analysis of vowel spectra", J. Acoust. Soc. Am. 41, 707-712.

Pols, L.C.W. (1977). "Spectral analysis and identification of Dutch vowels in monosyllabic words," Doct. Dissertation (Free University, Amsterdam).

Pols, L.C.W., Kamp, L.J.Th. van der, and Plomp, R. (1969). "Perceptual and physical space of vowel sounds," J. Acoust.Soc.Am. 46, 458-467.

Pols, L.C.W., Tromp, H.R.C., and Plomp, R. (1973). "Frequency analysis of Dutch vowels from 50 male speakers," J.Acoust.Soc.Am 53, 1093-1101.

Risset, J.C., and Wessel, D.L. (1982). "Exploration of timbre by analysis and synthesis," in The psychology of music, edited by D. Deutsch (Academic, London).

Rothenberg, M. (1982). "Acoustic interaction between the glottal source and the vocal tract," in Vocal Fold Physiology, edited by K.N. Stevens and M. Hirano (University of Tokyo, Tokyo).

Rothenberg, M. (1983). "An interactive model for the voice source," in Vocal Fold Physiology, edited by D.M. Bless and J.H. Abbs (College Hill, San Diego).

Russo, V., and Large, J. (1978). "Psychoacoustic study of the Bel Canto model for register equalisation: male chest and falsetto," J. of Research in Singing, 1(3), 1-25.

Rzevkin, S.N. (1956). "Certain results of the analysis of a singer's voice," Soviet Physics Acoustics, 2, 215-220.

Sambur, M.R. (1975). "Selection of acoustic features for speaker identification," Trans. IEEE-ASSP, 23, 176-182.

Schönemann, P.H., and Carroll, R.H. (1970). "Fitting one matrix to another under choise of a central dilation and a rigid motion," Psychometrica 35, 245-255.

Schouten, J.F. (1968). "The perception of timbre," Reports of the 6th ICA, Tokyo, GP-6-2, 35-44.

Schultz-Coulon, H.J., Battmer, R.D., and Riechers, H. (1979). "Der 3 kHz-Formant, ein Mass für die Tragfähigkeit der Stimme? II. Die trainierte Stimme," Fol. Phoniatr. 31, 302-313.

Smith, L.A., and Scott, B.L. (1979). "On increasing the intelligibility of sung vowels," J. Acoust. Soc. Am. 65, S124.

Smith, L.A., and Scott, B.L. (1980). "Increasing the intelligibility of sung vowels," J. Acoust. Soc. Am., 67, 1795-1797.

Sonninen, A., and Damsté, P.H. (1971). "An international terminology in the field of logopedics and phoniatrics," Folia Phoniatrica 23, 1-32.

Stumpf, C. (1890). "Tonpsychology," Vol 2 (Hirzel, Leipzig).

Sundberg, J. (1970). "Formant structure and articulation of spoken and sung vowels," Fol. Phoniatr. 22, 28-48.

Sundberg, J. (1973). "The source spectrum in professional singing," Fol. Phoniatr. 25, 71-90.

Sundberg, J. (1974). "Articulatory interpretation of the "singing formant"," J. Acoust. Soc. Am. 55, 838-844.

Sundberg, J. (1975). "Formant technique in a professional female singer", Acustica 32, 89-96.

Sundberg, J. (1977). "Vibrato and vowel identification," Archives of Acoustics 2, 257-266.

Sundberg, J. (1981). "Formants and fundamental frequency control in singing. An experimental study of coupling between vocal tract and voice source," Acustica 49, 47-54.

Sundberg, J. (1982). "The perception of singing," in The psychology of music, edited by D. Deutsch (Academic, London).

Sundberg, J., and Askenfelt, A. (1983). "Larynx height and voice source: a relationship?," in Vocal Fold Physiology, edited by D.M. Bless and J.H. Abbs (College Hill, San Diego).

Sundberg, J. and Gauffin, J. (1979). "Waveform and spectrum of the glottal voice source," in Frontiers of speech communication research edited by B. Lindblom and S. Ohman (Academic, London).

Van den Berg, Jw. (1955). "On the role of the laryngeal ventricle in voice production," Folia Phoniatrica 7, 57-69.

Van den Berg, Jw., and Vennard, W. (1959). "Towards an objective vocabulary for voice pedagogy," NATS Bulletin, February.

Vennard, W. (1967). "Singing, the mechanism and the technic," (Fisher, New York).

Von Bismarck, G. (1974a). "Timbre of steady sounds: A factorial investigation of its verbal attributes," Acustica 30, 146-159.

Von Bismarck, G. (1974b). "Sharpness as an attribute of the timbre of steady sounds," Acustica 30, 159-172.

Winckel, F. (1953). "Physikalische Kriterien für objective Stimmbeurteilung," Fol. Phoniatr. 5, 231-252.

Yilmaz, H. (1967). "A theory of speech perception," Bull. Math. Biophysics 29, 793-824.

Zahorian, S.A., and Rothenberg, M. (1981). "Principal-components analysis for low-redundancy encoding of speech spectra," J. Acoust. Soc. Am. 69, 832-845.

STELLINGEN

1. De wijdverbreide gedachte dat de bij 'kopstem' en 'borststem' in kop en borst waargenomen trillingen zouden bijdragen tot de geluidvorming is onjuist en bemoeilijkt de discussie over zangtechniek.

2. Het is onjuist om de overgang van stemregister bij een zanger te koppelen aan een bepaalde toonhoogte omdat de overgang sterk afhankelijk is van de geluidsintensiteit.

3. Er zijn tot op heden geen objektieve maten ontwikkeld die een onderscheid in meer dan zeven stemtypen rechtvaardigen.

4. De zangersformant is noch de enige noch een noodzakelijke voorwaarde voor een goede stemkwaliteit.

5. Het effekt van zaalakoestiek op zang is nog onvoldoende bekend; deze kennis is noodzakelijk voor begrip van perceptief belangrijke eigenschappen van de zangstem.

6. Een fonetogram, inklusief akoestische stemkwaliteitskenmerken, kan in de toekomst voor de foniatrie worden wat nu het audiogram is voor de audiologie.

7. De beschrijving van klankkleur kan vaak niets anders zijn dan een allerindividueelste expressie van een allerindividueelste emotie waarover men moeilijk van gedachten kan wisselen.

8. Het door Terbeek gepostuleerde niet-Euclidische karakter van de perifere auditieve perceptie is overbodig. Zijn resultaten laten zich even goed interpreteren via de niet-lineaire transformatie van formantfrekwenties in een Euclidische perceptieve ruimte.

Terbeek, D., UCLA working papers in phonetics 37 (1977).

9. Het geringe aantal publikaties over de vrouwenstem duidt meer op de complexiteit van de vrouwenstem dan op een ongeëmancipeerde instelling van veelal mannelijke onderzoekers.

10. Afgestudeerden in de fonetiek maken alleen een goede kans op de arbeidsmarkt als ze ook een grondige kennis hebben van spraaktechnologie.

11. Gezien de voortschrijdende luchtverontreiniging moet gevreesd worden dat er een tijd zal komen waarin de 'eeuwig zingende' bossen nog slechts gekreun zullen laten horen.