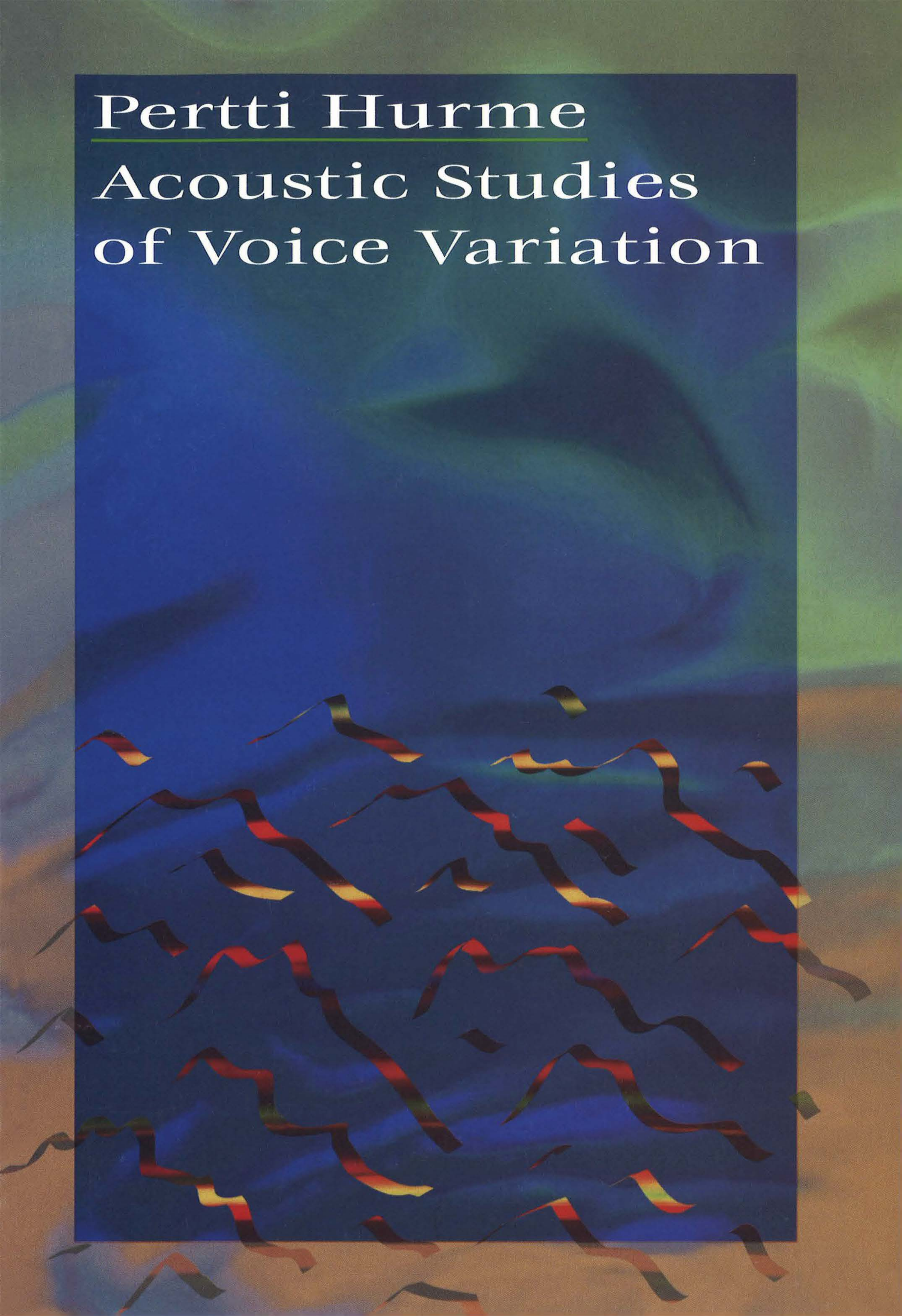
Pertti Hurme

Acoustic Studies of Voice Variation



Acoustic Studies of Voice Variation

JYVÄSKYLÄ STUDIES IN COMMUNICATION 7

Pertti Hurme

Acoustic Studies of Voice Variation

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 1996

Editor
Raimo Salokangas
Department of Communication, University of Jyväskylä
Technical editor
Kaarina Nieminen
Scientific Publishing, University of Jyväskylä

URN:ISBN:978-951-39-8089-4
ISBN 978-951-39-8089-4 (PDF)
ISSN 1238-2183

ISBN 951-34-0860-4
ISSN 1238-2183

Cover
Jorma Luotio

Kopi-Jyvä Oy and Sisäsuomi Oy, Jyväskylä 1996

ABSTRACT

Hurme, Pertti
Acoustic Studies of Voice Variation
Jyväskylä: University of Jyväskylä, 1996, 148 p.
(Studies in Communication,
ISSN 1238-2183; 7)
ISBN 951-34-0860-4
Finnish summary
Diss.

The present study deals with variation in the human voice. The functions of the voice are addressed and the major practical and theoretical approaches to the voice introduced. A three-fold approach is taken involving phonetics, logopedics and phoniatrics, and singing pedagogy. Voice research is described from three perspectives: voice production, acoustics and perception. A central concept is the adduction–abduction continuum, to which various descriptive categories of voice (e.g. creaky and breathy) are related. The study aims at describing the spectral properties of dysphonic voices, of voices varying in vocal intensity, and of supported and covered singing voices. The study also aims at relating the spectral observations to voice production, especially to the adduction–abduction continuum. In addition, the instrumental method used, long-term average spectrum (LTAS) analysis, is evaluated. Four sets of materials were investigated by means of LTAS analysis. Two were spoken materials: dysphonic voices (n = 87) and voices varying in vocal intensity (n = 10); two were sung materials: supported and unsupported voices (n = 8) and coverered and open voices (n = 1). The results indicate that the method used can differentiate between various voices. Several significant differences were observed in dysphonic voices, e.g. voices of patients with paralysis of vocal folds compared to those with laryngeal cancer. The spectra of loud and soft voices differed systematically. There were spectral differences in supported vs. unsupported voices, even though there was much individual and gender-related variation. The LTA spectra of covered and open voices differed systematically. Female and male voices also differed in spectral slope. Even though direct comparison of voices from the four data sets was not possible, the groups which showed distinct spectral differences appeared to form two clusters, ”powerless” and ”powerful”, characterized by steep and shallow spectral slopes, respectively. The study showed LTAS analysis, despite its limitations, to be a useful means of investigation into the human voice and its variation.

Keywords: voice variation, dysphonia, vocal intensity, supported voice, covered voice, gender, long-term average spectrum analysis

CONTENTS

PREFACE

PREFACE

The human voice is an enchanting object of study. It shows much variation, both within and between individuals. It can be investigated on many levels: as physiological and biomechanical activity, as an acoustic signal, and in the process of perception and interpretation in a social context. This study concentrates on the acoustic aspects of voice variation.

The roots of this study go back to the seminal work of Professor Kalevi Wiik at the University of Turku. I am grateful to him for introducing me to the fascinating world of phonetics and linguistics and also for providing me with the incentive to find my own path in the academic world. The path led to the University of Jyväskylä, where interdisciplinary research on voice started in the late 1970s. Collaboration with Professor Aatto Sonninen, the Grand Old Man of Finnish phoniatrics, has greatly widened my views about the human voice. I warmly thank him for the many scientific discussions and vigorous arguments we have had and continue to have while carrying out research and writing reports on various aspects of the human voice. I also thank Professor Jaakko Lehtonen for providing the impetus for my early studies of the human voice using long-term average spectrum analysis and for encouragement during the process leading to the completion of this dissertation.

I am grateful to the reviewers of the manuscript for this dissertation, Docent Britta Hammarberg and Professor Erkki Vilkman, for their insightful comments and suggestions that were useful in writing the final version.

My research activities have brought me into professional contact with many voice and speech researchers both in Finland and abroad. I have had the opportunity to spend shorter or longer study periods e.g. in Sweden, the United States, Italy and Russia. I wish to thank colleagues and friends, wherever they are, for the many fruitful discussions on a variety of topics related to the theme of this work. I also thank the staff and students at the Department of Communication where I work: I appreciate the open and future-oriented atmosphere which prevails at the department.

The financial support given me by the University of Jyväskylä and the Faculty of Humanities has made it possible for me to concentrate on writing this dissertation free from my teaching obligations during two semesters. I thank the university for accepting this work in the Studies in

Communication series and for giving me a grant toward the printing expenses. I also thank Professor Raimo Salokangas, the editor of Studies in Communication, and Kaarina Nieminen, the Secretary of Scientific Publishing, for their friendly co-operation. Thanks are also due to Graphical Designer Jorma Luotio for translating my ideas about the cover page into an artistic design and to Senior Lecturer Michael Freeman for professional help in revising my English.

My deepest thanks go to Maarit Valo, my colleague and my wife, for her unfailing support and encouragement. I particularly value our discussions, often at the kitchen table, on the human voice, communication and research.

I dedicate this work to my parents, Irma and Viekka, and to the next generation, Freija and Rauli. Research is the outcome both of experience and of experimentation.

Jyväskylä, November 1996

Pertti Hurme

1 INTRODUCTION

Human beings are skilful in modifying and coloring their voice to express themselves. Such modification can be temporary, as in expressing joy or fear. It can also be more permanent, as when adopted to indicate membership in a group. This capacity for variation in voice has been recognized by scientists for a long time. For instance, Aristotle reflected extensively on the subject. In On things heard (quoted in Barnes, 1984) he wrote:

> "It is the quickness of breathing which makes the voice shrill, force which makes it hard. So it happens that the same individuals have not only sometimes a shriller and at other times a deeper voice, but also at times a harder and at times a softer voice."

Aristotle was observing variation in a person's voice. According to him, the reasons for this variation are quickness of breathing and force. Quickness of breathing may refer to a relatively high air-flow through the glottis during voice production (phonation). Force, in turn, may refer to a relatively forceful activity of the vocal folds during voice production. Interpreted this way, Aristotle captured an essential aspect of voice variation, even though his terms differ from those used today.

The present study focuses on the human voice, more specifically on variation in voice, both in speaking and singing tasks. The general goal is to describe voice variation in the performance of different tasks and across groups of people, by means of acoustic instruments and measurements. Being noninvasive, acoustic methods permit the examination of voice produced in relatively natural circumstances, while yielding objective, quantitative data about voice characteristics.

This study is a reanalysis and reinterpretation of data and results published in several articles (see Appendix). The study also includes previously unpublished data. The articles on which the study is based have been published during a time span of more than fifteen years. The research reported in them is the result of acoustic analyses of voice carried out using the methods available at that time. Hence, the advances in research techniques are necessarily reflected in the methods employed.

The aim of this study is to describe voice variation by means of acoustic methods and relate that variation to a number of background variables and to voice production. To achieve this end, several perspectives and traditions in the study of voice will be reviewed. The four types of materials investigated represent major aspects of voice variation: (1) disorders in the speaking voice are studied by comparing groups formed on the basis of medical diagnosis; (2) normal speaking voice is studied in three loudness conditions, viz. loud, normal and soft voice; (3) the singing voice is examined by comparing supported and unsupported as well as covered and open singing; and (4) female and male voices are compared both in the speaking and singing tasks throughout the study. A subsidiary aim of this study is to evaluate the acoustic methods of analysis used.

The present study is a contribution to the acoustics of voice. It is motivated by a desire to know more about the human voice and about the factors underlying variation in voice. Such knowledge is important in itself, as is any knowledge about human beings. An insight into variation in human voice may contribute both to the theory of voice and to practical work in such areas as voice training, phoniatrics and singing pedagogy.

2 VOICE AND VOICE VARIATION

2.1 Definition of voice

The classification of the sounds that human beings can produce is a starting point for the definition of voice. Sounds can be produced with the vocal folds vibrating (voiced) or not vibrating (unvoiced). Unvoiced sounds are not the primary concern here, even though voices often contain unvoiced components. Voiced sounds or voice sounds can be divided into three categories (cf. Sundberg 1987, Titze 1994a):

1. Vocalizations. Vocalizations are sounds or sequences of sounds that have no linguistic structure. They typically express emotions or, in more general terms, physiological and mental states, by means of groans, sighs and cries, among others (expressing pleasure, fear, surprise, warning, pain, anger, joy etc.).
2. Singing sounds. In order to sing, the frequency of vibration of the vocal folds has to be controlled rather precisely. Typically singing also constrains the loudness and quality of voice. Even though we usually sing with words, it is possible to sing without words, humming a tune or vocalizing (in the musical sense of the word: singing with voiced sounds or syllables). Vocal music is made up of singing sounds.
3. Speech sounds. In order to speak, a number of articulatory maneuvers have to be executed and co-ordinated in time, guided by language. Simplifying somewhat, it can be said that speech is made up of speech sounds and language.

Vocalizations, singing sounds and speech sounds are short-term phenomena. However, they can be repeated or prolonged, so that a characteristic voice emerges. By voice is usually meant the relatively long-term (permanent or quasi-permanent) movements and positions of the organs of speech when speaking or singing – and the characteristic auditory coloring or timbre that results. Voice can be studied without language (vocalizations and musical vocalizations) and with language (singing with words and speech). Both the long-term and the short-term characteristics are produced by the same anatomical structures, which consequently can be called both organs of voice and organs of speech (or organs of singing).

Definitions of voice span a continuum from narrow to broad. A narrow definition regards voice as a laryngeal phenomenon, as regular vibrations of the vocal folds. However, the vibrations as such cannot be heard; in vocalization, singing and speaking the voice produced at the vocal folds is modified by the vocal tract and radiated from the mouth. What we hear is the acoustic signal. A broad definition regards voice as both a laryngeal and a supralaryngeal phenomenon. To take an example, a voice can be characterized by leakiness, due to an incomplete closure of the vocal folds during vibration, and by nasality, due to an incomplete closure of the nasopharynx by the velum. The former is a laryngeal phenomenon, the latter supralaryngeal. Other common terms for these phenomena are phonatory and articulatory: in a broad definition voice results from both phonation and articulation.

The broad definition is adopted in this study. The broad definition is adopted also in every-day language: one person has a pleasant voice, another an unpleasant voice, yet another a hoarse and rasping voice. Every-day characterizations are probably auditory, although self-perception may also connect with voice production.

In the study of voice, five levels of observation can be distinguished (see e.g. Laver 1980, Lehtonen & Hurme 1980, Scherer 1982, Hurme & Sonninen 1985a, Haskell 1987, Sonninen & Hurme 1992):

1 Anatomical constraints in the voice and speech production system (e.g. nodules in the vocal folds, cleft palate, the length of the vocal tract). These organic factors define a range of possibilities for voice production.

2 Long-term habitual adjustments of the voice and speech production system (e.g. habitual high pitch, habitual breathy voice, habitual nasalization).

3 Subjective impression of one's own voice. A voice does not sound the same to a listener as to its owner. This has to be taken into account when attempting to improve one's speaking or singing voice.

4 The acoustic speech signal, with its four basic parameters: duration, amplitude, fundamental frequency and spectrum.

5 Subjective impression of another person's voice, of the auditory colouring characteristic of a speaker.

Levels 1 and 2 pertain to the production of voice, levels 3 and 5 to the perception of voice and level 4 to the acoustics of voice.

Voice can be divided into different components, depending on the perspective adopted. From the perspective of production, voice consists of (1) the fundamental frequency of vocal fold vibration, (2) vocal intensity or amplitude of vocal fold vibration and (3) phonatory and articulatory properties. From an acoustic perspective, voice consists of (1) fundamental frequency, (2) amplitude and (3) spectrum. From the perspective of perception, voice consists of (1) pitch, (2) loudness and (3) voice quality or timbre. It is from these perspectives that voice is described below.

This study excludes properties that sometimes are included in voice. For instance, temporal factors such as speech tempo and fluency are not considered here. This decision is in accordance with Scherer (1979), among others: only the three basic properties are included in what he calls the vocal aspects of speech or vocal cues and what can simply be called voice as well.

In sum, voice is a convenient term. It covers the three sets of components presented above. More technical terms can be used when appropriate, to emphasize a certain aspect of voice, e.g. voice quality, when perceptual aspects of voice are emphasized.

This study encompasses many kinds of voices: speaking voice and singing voice, dysphonic voice and normal voice, good voice and poor voice, female voice and male voice. Therefore, voice and voice variation will be described extensively below. The emphasis is on phonatory and spectral properties, but other aspects will be discussed too.

2.2 Evolution and ethology of the human voice

In the prehistory of mankind, singing may have had a role in the progress from vocalizations or prespeech to speech. The use of the auditory-vocal channel for survival (e.g. warnings and threats) is common in vertebrates. It is efficient for inter-species communication in the absence of eye contact, e.g. in the dark. In the process of evolution the ear preceded the vocal organs – constraining the evolution of the vocal organs (as the production of sounds that could be received by the ear was required).

Like other primates, the ancestors of humans must have been able to produce vocalizations for millions of years. These vocalizations have gradually developed into more complex structures (see e.g. Damasio & Damasio 1992). Several researchers have proposed that singing may have acted as a link between vocalizations and speech (see e.g. Luchsinger & Arnold 1965, Levman 1992, Richman 1993). Korhonen (1993) even plays with the idea that hominids may have used their voice in courting behaviour, resulting in the separation of vocalization from its original context, a process sometimes called ritualization. If vocalizations were used in courting, it is plausible that such vocalizations tended to become more beautiful and more impressive. In other species, analogies for such behaviour abound, e.g. the song of the nightingale. It is probable that singing has been a human characteristic for a long time, possibly before the emergence of speech and language. The human voice, including vocalizations (e.g. cries of pain) and singing, can evoke strong emotions in other human beings.

Ohala (1984) has examined the utilization of voice from an ethological perspective. Ohala associates low pitch with the desire to appear large and threatening, the intention to triumph in a contest of power, dominance and self-sufficiency and high pitch with the desire to appear small and unthreatening as well as with submission, appeasement and the desire for goodwill and co-operation. The sounds made by a confident aggressor (or one who wants to appear so) are typically rough and have a low pitch; submissive and non-threatening individuals' cries are typically tone-like and have a high pitch. This "frequency code" is not only based on pitch but also on timbre, regulated by means of vocal tract resonances. The vocal tract can be made larger by protruding the lips and/or by lowering the larynx; thus, lower resonances (formants) are produced. The vocal tract can be made smaller by spreading the lips and/or raising the larynx; thus, higher resonances (formants) are produced. In animals, according to Ohala (1984), lip protrusion sometimes accompanies threatening behaviour, lip spreading submissive behaviour. Ohala extends the idea of the frequency code to gender differences in human beings. According to him, sexual dimorphism has been observed in the vertical larynx position (VLP): male VLP is lower than female, making the vocal tract larger and the resonances lower (thereby accentuating gender differences in voice).

Ohala (1984) discusses the frequency code mainly from the perspective of pitch (fundamental frequency). His ideas can, however, be extended to two other basic properties of voice: loudness and timbre. It is plausible that a loud voice indicates a big animal (or an animal that wants to appear big), a soft voice a small animal (or an animal that wants to appear small). Titze (1994a) maintains that the entire animal world seems to respect high levels of acoustic power. Ohala (1984) associates low pitch with a rough voice and high pitch with a tone-like voice, as described above. The spectral properties of voice will be discussed in detail later. Suffice it here to point out that a rough voice may be related to rich timbre (due to rich overtones) and a tone-like voice to poor timbre (due to weak overtones). Hypothetically, low pitch, loud voice and "overtone richness" would characterize a big animal (or an animal that wants to appear so), and high pitch, soft voice and "overtone poorness" a small animal (or an animal that wants to appear so).

Animal behavior gives insights into human communication. Even though human communication is more complex than animal communication, humans and animals alike share a basic building block of communication: the voice and its potential for expression.

2.3 Functions of voice

Much can be heard in a voice: attitude to the listener, attitude to what the speaker is saying, emotional state, condition of health, social group, dialect background, sex, age etc. (see e.g. Laver & Trudgill 1979, Brown & Bradshaw 1985, Pittam 1994).

Even though Quintilian (quoted in Laver 1980:1) wrote two millennia ago that the voice of a person is as easily distinguished by the ear as the face by the eye, voice is not an entirely stable cue to identity. Human beings are good at imitating other persons (see e.g. Kent 1973, Hurme 1976). A person's voice varies across speech situations. Inadvertent changes may be caused by illness and fatigue. Nevertheless, there is often a sufficiently constant core of acoustic similarity (invariance) to allow listeners to make decisions about identity.

Voice also gives an impression of the speaker. The impression can be true, e.g. the voice may indicate that the speaker is tired or that he or she has caught a cold. The impression can also be untrue, based on the listener's beliefs and stereotypes. For instance, negative qualities may be assigned to the pitch of Western women speaking at their normal pitch level in the Japanese speech culture, where women's pitch is expected to be higher (Yamazawa & Hollien 1992, van Bezooijen 1995). Impressions of a person from vocal cues have been investigated extensively (e.g. Scherer 1978, Brown & Bradshaw 1985, Pittam & Gallois 1986, Valo 1994).

Voice has many functions. A primitive use of voice is for survival (e.g. warning signals, expression of hunger). The expression of emotion is also highly vocal. A person's voice is very much a part of his or her personality. It conveys information, for example, about the individual, about his or her attitudes, and about his or her affiliation to groups. The voice also conveys information about a person's health. Last but not least, the human voice "carries" both singing and speech.

There are three major groups of indices or markers in linguistic analyses of voice or, to use a more technical term, long-term speech characteristics (see Abercrombie 1967, Laver & Trudgill 1979):

1 Social markers: regional affiliation, social status, educational status, occupation, social role.

2 Physical markers: age, sex, physique, state of health.

3 Psychological markers: characteristics of personality, affective state.

Social markers give information about a person as a member of a community. All languages have regional and social dialects with distinct voice quality settings that function as indicators of membership in a group. Voice can indicate social class, but the markers are not the same for different variants of the language. For instance, a higher incidence of creaky voice in the higher-class males of an Edinburgh sample was reported by Esling (1978), whereas Trudgill (1974) found a higher incidence in working-class speakers in Norwich (see also Pittam 1987, Henton & Bladon 1988, Esling, Harmegnies & Delplancq 1991). Excessive nasality indicates submission in some cultures and power in other cultures (Laver & Trudgill 1979). Regional affiliation also shows in voice (Esling 1982, Elert & Hammarberg 1991).

Differences in voice across languages have been sought and to some extent also found (e.g. Tarnóczy & Fant 1964, Majewski, Rothman & Hollien 1977, Bruyninckx, Harmegnies, Llisterri & Poch-Olivé 1994), but in such studies methodological problems abound: how can comparable samples of speech be obtained from speakers of different languages (with widely differing distributions of vowels and consonants, for instance)? In a recent and well-controlled study no differences between languages were found (Byrne, Dillon, Tran, Arlinger, Wilbraham, Cox, Hagerman, Hetu, Kei, Lui, Kiessling, Kotby, Nasser, El Kholy, Nakanishi, Oyer, Powell, Stephens, Meredith, Sirimanna, Tavartkiladze, Frolenkov, Westerman & Ludvigsen 1994).

Occupation and social role mark many voices. For instance, priests and ministers sometimes adopt a clerical voice. Much is expected of actors' and singers' voices in terms of aesthetic quality. Such aesthetic qualities, however, vary widely in different cultures. Singers develop voice qualities prevalent in their music culture (for instance, Western operatic singing, Korean pansori singing, see e.g. Sonninen 1987, Sundberg 1987, Sataloff 1991, Leino 1994). A major criterion, however, for the voice on the stage – both spoken and sung – is that it needs to be loud enough to be audible in the auditorium.

Physical markers include age, physique, sex and state of health. The size of the vocal folds and the vocal tract obviously has some effect on voice. An infant has a higher pitch and higher formants than an adult. However, the effect is far form straightforward: vocal fold and vocal tract anatomy only determines the constraints defining the range of possibilities for varying the voice. An individual has wide possibilities for changing his or her voice in terms of pitch, loudness and quality. Singers have to be capable of extreme maneuvers. The voice class of singers usually goes with the size of the person (see e.g. Dmitriev & Kiselev 1979, Seidner & Wendler 1982).

Female and male differences in the fundamental frequency (F0) depend partly on anatomical differences, but also on learned differences (see e.g. Ohala 1981). Differences in pitch may be accompanied by differences in vocal quality, e.g. by breathiness in the female voice and by roughness in the male voice. Even though female and male differences are included among the physical markers here, they are clearly also social markers: being a woman or a man is, in addition to the physical body, a social role (see e.g. Henton & Bladon 1985).

The idea of voice as a symptom of a disease or disturbance has lead to many attempts at screening for the detection of illness on the basis of voice. For instance, Ostwald (1973) thought he had found the voice of psychosis from acoustic measurements. However, it is highly improbable that there is an invariant "psychotic" voice in a disturbance manifesting itself in such a variety of ways. Attempts to associate voice properties with mental phenomena often assume too simple a connection between a measurable acoustic phenomenon and its cause.

Many attempts have been made to reach a more realistic goal, to detect a laryngeal pathology (e.g. cancer) on the basis of acoustic screening (e.g. Kasuya, Masubuchi, Ebihara & Yoshida 1986, Laver, Hiller, MacKenzie & Rooney 1986). Such attempts may ultimately prove to be useful. The connection with the medical state of a person is obvious in many types of voice disorders (e.g. those resulting from physiological malfunctioning or breakdown).

Psychological markers include characteristics of personality and affective state. These can perhaps less clearly be regarded as long-term characteristics; for instance, affective states are usually rather transitory, medium-term in the Laver & Trudgill (1979) framework to be described below. There is a long line of research in this area (e.g. Williams & Stevens 1972, Scherer 1986, Baltaxe 1991, Pittam 1994). Nevertheless, there is relatively little systematic knowledge of the details of the encoding and decoding processes of psychological markers (Pittam & Scherer 1993).

2.4 Approaches to voice

Pittam has recently summarized the principles underlying the study of voice (1994:123) as follows:

> 1 Voice is a communicative channel comprising a package of nonverbal (specifically, vocal) behaviors.
>
> 2 Like other nonverbal channels, voice may be used in social interactions to communicate group and personal identity, affect, attitude, and so forth.
>
> 3 Voice is used with speech in most interactions; it underlies all speech.
>
> 4 Voice can be described using the same types of articulatory and acoustic parameters as those used for speech.

The first three points summarize the discussion presented above. The fourth point argues that voice can be described in terms of the parameters used in the description of speech. The present study has its closest connections with three traditions of speech and voice research: the phonetic tradition, the phoniatric-logopedic tradition and the singing pedagogy tradition. They all have arisen from practical needs: the desire to help people learning foreign languages, to help people suffering from voice disorders, and to help people sing better. All three traditions are ancient, dating back centuries or millennia (Panconcelli-Calzia 1961, Daoud 1965, Laver 1981, Stemple 1984, Perelló 1987).

2.4.1 Phonetic tradition

Phonetics is concerned with describing how we, speakers of more than 6000 languages, speak and understand speech. For this purpose, many studies on the production, acoustics and perception of speech sounds have been conducted, particularly during this century, but also earlier. In recent decades this enterprise has more and more become to involve various suprasegmental phenomena, including voice quality (see e.g. Laver 1994, Ladefoged & Maddieson 1996).

In the phonetic tradition, voice production is typically described in the following manner: voice is produced in the vocal organs or vocal apparatus. The necessary energy (in the form of subglottal pressure) is created in the respiratory system, whence it is converted into acoustic energy in the phonatory system and (in the form of a glottal pulse) modified by the articulatory system into voice. In other words, subglottal pressure is translated into a sound wave at the glottis by the vibration of the vocal folds and modified in the vocal tract.

This process is described in Figure 1 (adapted from Sundberg 1987). The processes leading to the production of speech are described with regard to the organs and the muscle groups involved. Even though the phonetic tradition grew out of the study of the speaking voice, it has been extended to the singing voice as well. The figure also shows the three functions

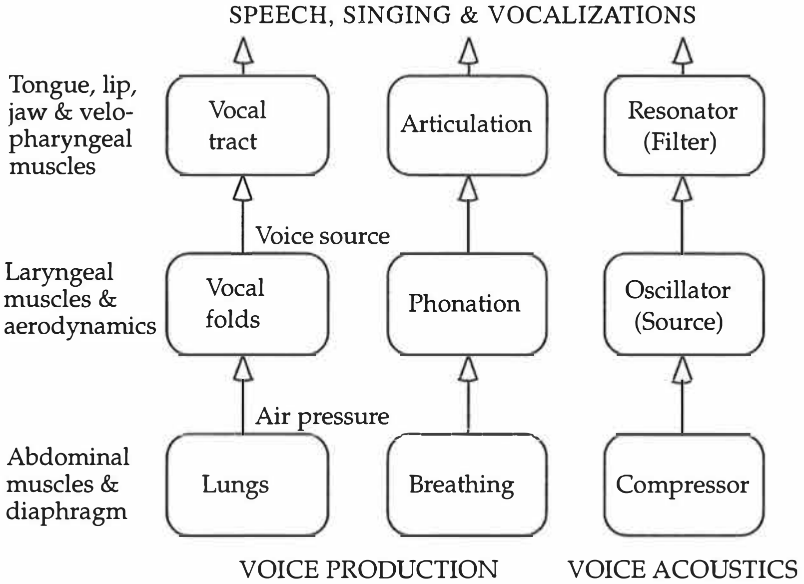


FIGURE 1 Voice production related to the vocal organs and to their physiological and acoustic function (adapted from Sundberg 1987).

common in the description of speech production: breathing (respiration), phonation and articulation. In addition, the corresponding acoustic processes are shown.

From the point of view of phonation and articulation, voice is composed of respiratory, phonatory and articulatory movements, resulting in variation in the fundamental frequency of phonation, in vocal intensity and in vocal tract characteristics.

Figure 1 also describes the acoustics of voice and speech production. From an acoustic point of view, speech and voice production consists of a compressor, an oscillator and a resonator. The compressor builds up pressure, the oscillator produces an acoustic signal and the resonator modifies the signal. When speaking, the lungs build up subglottal pressure and an airstream is sent up the trachea. If the vocal folds are in a favourable position (adducted), they start vibrating, i.e. opening and closing the glottis, chopping thereby the airstream into a series of air pulses. These pulses raise the air pressure above the glottis at regular intervals. In speech and voice production, the oscillator is the glottal source; the resonator acts like a filter, boosting some frequencies and damping others, depending on the properties of the resonator. Therefore, what has been outlined above is often called the source–filter theory of speech/voice production. There are a number of studies (e.g. Fant 1960, Stevens 1977) and textbooks (e.g. Fry 1979, Kent & Read 1992) on the acoustics of speech and voice.

The chain of events can be more technically described as consisting of four phases:

1 The glottal source waveform describes transglottal air flow across time.

2 The glottal source spectrum describes the amplitude of the harmonics (fundamental frequency and its overtones). The properties of the source waveform determine the spectral slope of the glottal source spectrum.

3 The vocal tract transfer curve determines the locations of resonances (formants) on the frequency scale; in other words, the glottal spectrum is filtered according to the transfer characteristics of the vocal tract.

4 The spectrum radiated from the lip opening is the result of the previous processes; the glottal source spectrum is the input to the vocal tract resonator, and the radiated spectrum is the output.

The glottal source spectrum gives information about the voice source, the radiated spectrum about the entire process. There are methods for studying each of these.

Settings

To describe phonation and articulation, the phonetic tradition has a useful concept, the setting. In the words of Abercrombie (1967:93), an articulatory setting is "so deeply rooted as to seem as much an unalterable part of a person as his anatomical characteristics". The concept has its origin in the centuries-old organic base or basis of articulation (see e.g. Sweet 1877, Laver 1981), which has been used to classify and evaluate languages and speakers of these languages as having an anterior or a posterior basis of articulation. An anterior basis of articulation is typically associated with positive connotations and a posterior with negative. Basis of articulation is a theoretically diffuse cover term, which clearly reflects the prestige attributed to speakers of certain languages (see e.g. Kelz 1971, Laver 1978).

Honikman (1964:73) built on the tradition of the organic base. According to her, an articulatory setting is "the fundamental groundwork which pervades and, to an extent, determines the phonetic character and specific timbre of a language". However, Honikman also attempted to break down the concept into its components. For instance, in a comparison of English and French, she observed differences in lip movements: in English the lips are neutral and moderately active, in French rounded and vigorously active in spreading and rounding.

In a classic treatise, Abercrombie (1967:89) distinguished three strands, "separable though closely woven together, all simultaneously and continuously present and together making up the [aural] medium". The strand consisting of (1) the segmental features of an utterance is made up of complex auditory qualities in fairly rapid fluctuation (reflecting the rapid movements of the articulators). The strand of (2) voice dynamics consists of considerably slower fluctuations in auditory quality. The features of voice dynamics are closely related to those aspects of sound that assume importance in music – such as pitch, tempo, loudness and rhythm. In contrast to the first two, the strand consisting of (3) voice quality has a quasi-permanent character. Voice quality remains constant over long stretches of time, with much less apparent fluctuation. The three strands can be described in a time perspective as short-term, medium-term and long-term. Abercrombie (1967:90) gives the following examples of the three strands: Educated voice and clear voice refer to the strand of segmental features. Loud voice and low voice (low either in pitch or low in volume) refer to voice dynamics. Pleasant voice and hoarse voice are connected with features of voice quality. The examples given by Abercrombie may be illustrative, but on closer inspection they are somewhat ambiguous. For instance, loud voice can also be a long-term characteristic of a person's voice. In fact, all these observations apply equally to other strands (Hurme 1985).

Laver and Trudgill (1979) have developed a theoretical framework for the description of voice. Their taxonomy is reproduced in Table 1: voices are classified on basis of their signalling function, their relation to language, temporal perspective, vocal variables, marking function and potential controllability.

TABLE 1 Essential classifications in the Laver & Trudgill (1979) framework for the description of voice.

VOCAL VARIABLES	Vocal features deriving from anatomical differences between individuals	Voice settings, i.e. habitual muscular adjustments of the vocal apparatus	Paralinguistic 'tone of voice' achieved by temporary use of voice settings	Momentary articulatory realizations of phonological units
TEMPORAL PERSPECTIVE	Permanent	Quasi-permanent	Medium-term	Short-term
POTENTIAL CONTROLL-ABILITY	Uncontrollable, therefore unlearnable	Under potential muscular control, therefore learnable and imitable		

A certain vocal quality may have many functions. For instance, a rough or tense voice may result from a malfunction of the vocal folds due to a vocal nodule, from the adoption of a habitual rough setting to indicate membership in a group, from temporary aggressiveness or anger, or from a phonetic distinction in a language with distinctive voice qualities (see e.g. Ladefoged & Maddieson 1996). The framework gives a solid foundation onto which vocal phenomena can be mapped.

In the phonetic tradition, much effort has been given to constructing a scientific terminology, with special emphasis on articulatory terms. The majority of the terms probably apply to short-term, segmental phenomena, but terms and descriptive systems have also been developed for longer-term phenomena, e.g. by Catford (1964, 1977), Laver (1980, 1994) and Ladefoged and Maddieson (1996). Their work has its foundation in centuries of work aiming at understanding human speech and at developing a scientific terminology in the area.

2.4.2 Phoniatric-logopedic tradition

Phoniatrics and logopedics are concerned with voice disorders, and with speech, language and communication disorders. The starting-point is often clinical, as the examination of voice aims at detecting the cause of the disturbance (e.g. vocal nodules, paralysis of the recurrent nerve, velopharyngeal insufficiency) and at alleviating the problem. In the examination of voice, both articulatory, acoustic and perceptual methods are used (see e.g. Fritzell 1973, Hirano 1981a, Stemple 1984, Sataloff 1991).

While the phonetic tradition is, among other things, interested in the description of normal voice, in all its variation in the languages of the world, the phoniatric-logopedic tradition is interested in the description of the wide variety of disordered or dysphonic or pathological voices. A voice disorder is often defined by contrasting it to the voice of other persons of similar age, sex, cultural background, and geographic location: if a person's voice differs from those of other persons in quality, pitch and loudness to the degree that attention is drawn to him or her, a voice disorder is said to exist. A voice disorder may also be present when the structure or function of the laryngeal mechanism (or both) no longer meet the requirements established for the mechanism by the owner of the voice. Persons like singers and actors with a need for a good voice may be greatly concerned at even minor vocal difficulties. On the other hand, those with low vocal needs may fail notice the existence of vocal problems that would be considered severe by those with high vocal needs. (See e.g. Aronson 1980, Stemple 1984.)

More variation exists in disordered voices than in normal voices, as the causes of disturbances are highly varied. For instance, laryngeal cancer can increase adduction of the vocal folds, making the voice rough and harsh. Or to take another example, paralysis of the recurrent nerve may increase abduction of the vocal folds, making the voice weak and breathy.

In the phoniatric-logopedic tradition voices can be described by the etiology of the illness. Voices can also be described physiologically (e.g. by the amount of subglottal pressure or transglottal air flow, see Hirano 1989), acoustically (e.g. by spectral measurements, see e.g. Yanagihara 1967, Hirano 1981a) and perceptually (e.g. by professional raters trained in a system of voice classification, see e.g. Isshiki, Okamura, Tanabe & Morimoto 1969, Laver, Wirz, MacKenzie & Hiller 1981). Combinations of etiological, physiological, acoustic and perceptual approaches are common (see e.g. Hammarberg, Fritzell, Gauffin, Sundberg & Wedin 1980, Hurme & Sonninen 1985a, Hammarberg & Gauffin 1994).

Such studies are motivated by a desire to help individuals with dysphonia. They are also motivated by the possibility of constructing devices to help in the diagnosis of dysphonia or in the screening of risk groups. To be able to construct such devices, a minute description of dysphonic voices is needed.

2.4.3 Singing pedagogy and singing research tradition

The singing voice differs from the speaking voice. There are both phonatory and articulatory differences, as extensively described by Sundberg (1974, 1987) and others. Like the speaking voice, the singing voice shows variation. Ethnomusicological studies show that singing exists in many forms with widely differing voice quality ideals (Lomax 1978, Leistiö 1995). What is regarded as beautiful varies greatly across cultures. Standards and ideals have arisen. In these standards quality or timbre is crucial: it should fulfil the aesthetic expectations of the audience.

The rise of singing pedagogy in Western musical culture is presumably connected with the increasing demands set on the singing voice in the 19th century. Pure bel canto singing was no longer felt expressive enough – more variation was required of the singing voice. Singing pedagogues or voice teachers emerged to help singers meet the demands of audiences and composers (see Sonninen 1987).

The singing pedagogy tradition emphasizes the perceptual aspect of singing: how the singer hears and feels his or her voice and how the listeners hear the voice. Descriptive terms for various voice qualities abound in the practice of singers and singing teachers (Sonninen & Damsté 1971, Sonninen & Hurme 1992, Titze 1994b). For instance, the terms covered and open (dating back to about 1830), are used to describe certain varieties of good and poor singing voices.

The scientific study of singing owes much to singing pedagogues. They played an important role in the advancement of physiological studies of the voice. In 1854 Manuel García Jr., a celebrated singer and singing pedagogue, invented the laryngeal mirror with which the vocal folds could directly be observed (Perelló 1987). In more recent times, much research has been carried out on the physiology, acoustics and perception of the singing voice – to the point that a science of the singing voice is beginning to emerge (e.g. Sonninen 1987, Sundberg 1987, Sataloff 1991).

Singing pedagogy often relies on impressionistic labels for voices. The lack of a common terminology has been acknowledged – and lamented (see e.g. Mörner 1963, Michel 1985, Sundberg 1987). On the other hand, it is true that progress has been made in the scientific study of the singing voice, which started with García more than 150 years ago, to develop the terminology of singing pedagogy.

Two common descriptive terms are covering and support. Covered singing (gedeckte Stimme in German, voix couverte in French) was introduced as early as the beginning of the 19th century (Large 1972). This term is contrasted with open singing (offene Tongebung in German, voix blanche in French, see Luchsinger & Arnold 1965, Hertegård, Gauffin & Sundberg 1990, Miller & Schutte 1994). Another common term is support, also known as breath support (Italian appoggio, German Stütze, see Arnold 1973, Sonninen 1993, Titze 1994a, Griffin, Woo, Colton, Casper & Brewer 1995). Covered and supported singing will be described in chapters 2.5.2 and 3.2.3 in more detail.

The three traditions presented above demonstrate that the human voice can be approached from very different perspectives. As the three traditions have developed into separate disciplines, they also have developed partly diverging theories and terminologies. Nevertheless, they are united by a common object of study, the human voice.

2.5 Study of voice

2.5.1 Historical notes

The earliest systematic observations of the human voice were probably auditory, by ear. For instance, in the ancient physiognomic tradition the speaker and his or her voice were not distinguished. On the contrary, thorough-going inferences about a person's character were made on the basis of his or her voice. Such inferences are still sometimes made e.g. in music criticism: a singer's character may be judged from his or her voice.

There is a wealth of literature on the auditory characteristics of voice and voice quality, spanning more than two millennia (Laver 1981). Ancient terms are sometimes still used. For instance, the term orotund dates back to Horace (Laver 1981). However, the exact meaning of the term (from Latin ore rotundo) at that time appears unclear. The 19th century practice in vocal pedagogy may have equalled "la voix orotunde" and covered voice. In 19th century rhetorics the term orotund describes "an eminent degree of fulness, clearness and smoothness" (Rush, quoted in Laver 1981). Such terms may take on many meanings, as the connection with voice production is vague. On the other hand, certain very old terms such as nasal appear to have worked well in the past and still do. This may be due to the fact that this term is rather transparent: a nasal quality arises from "speaking through the nose".

The study of voice production is also very ancient. An Egyptian papyrus from about 1600 BC describing a voice disorder also includes a hieroglyph portraying the lungs and trachea. However, the larynx was not pictured in the hieroglyphs, as no organ of voice had yet been identified. (Stemple 1984.) In ancient Greece, Hippocrates acknowledged the importance to voice production of the air flowing from the lungs through the glottis. He saw loudness of voice as depending on the volume and intensity of the air-flow (Reich 1950). Hippocrates also stated that observation of voice quality, whether clear or hoarse, is one way in which to reach a diagnosis (Stemple 1984).

Aristotle was the first to mention that the larynx is the organ of voice. He also observed a connection between vocal pitch and the size of the vocal apparatus (Reich 1950, Barnes 1984). Nevertheless, the basic anatomy of the larynx was not described until the 2nd century AD by Galen of Pergamum (see Panconcelli-Calzia 1961). In experiments with pigs, the pigs stopped squealing when the recurrent laryngeal nerve was severed. This way Galen proved that the larynx was the organ of the voice and that the voice was controlled by the brain. Thus, Galen invalidated the belief popular at that time that the voice was sent forth by the heart. (Stemple 1984.)

Galen's description of the larynx lived on among Arab physicians (e.g. in the 10th and 11th centuries), who also made numerous observations on voice disturbances (Daoud 1965). The fact that the recurrent nerve is concerned with voice was also appreciated by the Arab physicians (see Panconcelli-Calzia 1961, Daoud 1965).

Nevertheless, in the Western world it took a long time for an appreciation of the vocal folds as the source of the voice to evolve. As late as the 17th century a scientific approach to voice production could be labelled as blasphemy: the voice was a gift from God and could not arise from anything as lowly as the lungs, the larynx and the mouth.

In 1700, Dodart claimed that the glottis is the principal organ of voice (Cooper 1989). Dodart's idea received support from experiments with excised larynges by Ferrain in 1744 (Panconcelli-Calzia 1961):

> "Le larynx du cadavre répondit par un éclat qui étonna les assistans, et c'est, je pense, la premiére fois qu'on a vû pareil phénomène."

This lowly source of voice slowly gained acceptance (see e.g. Sonninen 1987, Cooper 1989). It was not possible to observe the vibrations of the vocal folds in a living human being until the middle of the 19th century. The invention of the laryngeal mirror by García allowed direct observation of the vocal folds. In the latter half of the 19th century, the emerging sciences of physiology and phonetics contributed to a better understanding of voice physiology (Panconcelli-Calzia 1961, Perelló 1987).

This concise history of voice research clearly shows that what has since been regarded as self-evident may earlier have been clouded by misconception and even superstition. The human capability for self-perception of vocal phenomena is by no means self-evident: the function of the vocal folds was not understood until Dodart and Ferrein. Not everything is understood even now.

2.5.2 Production

As described above, the study of voice production, the anatomy of the vocal organs and the physiology of the vocal folds, has a long history. As the present study focuses on the spectral properties of voice, the contribution of both laryngeal and supralaryngeal structures will be considered.

Human beings make use of their voice in many ways. The individuality of voice depends on anatomy (primarily the vocal folds and the vocal tract) and on acquired speaking habits. Anatomy does not determine absolute acoustic values – rather it gives the ranges within which speakers can vary their voice. For instance, a speaker is within certain limits free to adopt a higher or a lower pitch, habitually or temporarily.

The frequency at which the vocal folds vibrate can be regulated by altering the length of the vocal folds, mainly by means of movements of the laryngeal cartilages caused by the cricothyroid muscle (Sonninen, Hurme &

Vilkman 1992, Titze 1994a). However, there are a number of other factors influencing the fundamental frequency of phonation. More subglottal pressure generally results in a higher pitch. There are also other muscles (external laryngeal muscles, e.g. Sonninen 1956, Vilkman, Sonninen, Hurme & Körkkö 1996) that participate in the regulation of the fundamental frequency. In all, the regulation of the frequency of vocal fold vibration is very complex. In addition to longer-term or "macro-level" variation in fundamental frequency (e.g. intonation in speech and notes in singing), there is also short-term or "micro-level" variation (e.g. jitter in the speaking voice and vibrato in the singing voice).

There are three potential mechanisms for controlling vocal intensity: below the larynx, at the larynx and above the larynx. Vocal intensity is mainly controlled by subglottal pressure (Gauffin & Sundberg 1989): the higher the pressure, the higher the vocal intensity. However, at the laryngeal level, there appears to be a glottal width at which intensity is maximized: both more adduction (resulting in a narrower glottal chink) and more abduction (resulting in a wider chink) reduce vocal intensity (Titze & Scherer 1987, Titze 1994a). Maximum intensity requires correct adductory positioning, a skill mastered by singers. Consequently, excessive vocal effort (hyperadduction) may not lead to a very loud voice. The third, supralaryngeal mechanism involves the adjustment of formants so that they increase the amplitude of one or more harmonics (see e.g. Carlsson & Sundberg 1992, Miller & Schutte 1994). As the facts of voice production have not been (and still are not) completely understood, other explanations have been sought for the differences observed in vocal intensity (and concomitant differences in quality). In voice pedagogy, one such attempt is the concept of support or breath support, to be described later in this chapter.

Voice is influenced by both laryngeal and supralaryngeal factors or settings (Honikman 1964, Laver 1980). Laryngeal settings will be described first, then supralaryngeal settings.

Three types of laryngeal settings have usually been distinguished in the phonetic tradition: voicing or voiced, unvoiced and whisper. Clearly, these three do not suffice to describe the full range of human phonatory capabilities. Summing up many studies (including their own) that have broadened the scope of phonetic research, Ladefoged and Maddieson (1996) examine the vocal fold vibration continuum from breathy voice to creaky voice: from the most open (abducted) setting of the vocal folds in which vibration will occur to the most constricted (adducted) setting in which vibration will occur. A setting more open than breathy voice is voiceless, a setting more constricted than creaky voice is glottal closure. Ladefoged and Maddieson distinguish the following laryngeal settings: glottal closure, creaky voice, stiff voice, modal voice, slack voice, breathy voice, voiceless. They emphasize that the categories chosen represent a continuum (Ladefoged and Maddieson 1996:49):

> "We have chosen to name only these seven major phonetic categories, which, generally speaking, will be sufficient to enable us to describe the surface phonetic contrasts that we have observed; but we would also emphasize that there is a continuum of glottal opening, and a different number of steps might have been named."

The classifications by other researchers (e.g. Catford 1964, Lindqvist Gauffin 1972, Stevens 1977) may differ in some points from those presented above. The differences are due to theoretical standpoints, to the range of voice qualities covered as well as to the actual terms used. Nevertheless, the settings describe a continuum of phonations centering around modal or normal voice. More adduction (and less air-flow) shifts the phonation toward tense, stiff, harsh or creaky voice, less adduction (and more air-flow) toward lax, slack, breathy or whispery voice.

In addition to phonatory settings and lax vs. tense overall tension, Laver (1994) distinguishes articulatory settings, over-all muscle tension and settings of articulatory range. Articulatory settings consist of longitudinal settings (which affect the length of the vocal tract), cross-sectional settings (labial, mandibular, lingual, and pharyngeal settings, which impose deviations from a neutral long-term configuration), and velopharyngeal settings (which impose deviations from the neutral non-nasal setting). Overall muscle tension can be lax or tense; lax vowels show a smaller degree of deformation of the vocal tract from the position of the vowel [ə] than tense vowels. However, the settings vary in their effect on the quality of voice. In a study of the acoustic consequences of various settings, Nolan (1983, see also Nolan & Kühnert 1995) concluded that phonatory settings have more effect on the acoustic signal than the other settings. Therefore, the main emphasis in this chapter is on phonatory settings.

Non-dysphonic voices tend to be close to modal (normal) voice, even though they may vary somewhat toward a more tense or creaky voice or toward a more lax or breathy voice. However, what is normal is far from a simple decision. Some speakers who regard their voice as normal might even be characterized as dysphonic in a phoniatric examination (cf. Ladefoged & Antoñanzas-Barroso 1985).

Dysphonic voices show more variation in the continuum from glottal closure to an open, voiceless state of the glottis. For instance, the development of a growth in the larynx due to cancer may result in a partial closure of the glottis, a paralysis of the vocal fold or folds may lead to aphonia. The continuum can be divided into seven categories, as in the Ladefoged & Maddieson (1996) system described above. However, the terms used in the logopedic-phoniatric literature often differ from those current in the phonetic-linguistic literature (e.g. Isshiki et al. 1969, Hurme 1986, Hammarberg & Gauffin 1994).

Table 2 summarizes some of the voice quality terms current in phonetic-linguistic and phoniatric-logopedic descriptions. The terms (which are merely examples of many terms in use) are related to the adduction–abduction of the vocal folds. The two extremes are closed glottis and open glottis. In-between is the normal or modal opening and closing movement of the vocal folds, which can be characterized as a balance of opening and closing forces; hence the abbreviation "balance" in the figure. The numbers in the adduction–abduction column indicate degrees of more adduction (positive values) and less adduction (i.e., more abduction, negative values). On the scale, plus or minus two represents more adduction or abduction than plus or minus one.

TABLE 2 Common descriptive terms in the phonetic-linguistic and phoniatric-logopedic traditions related to adduction–abduction of vocal folds.

Step	Adduction/ Abduction	Phonetic-linguistic description	Logopedic-phoniatric description
1	closed	glottal stop	glottal attack
2	+2	creaky, laryngealized	creaky, glottal fry
3	+1	stiff, tense	strained, pressed
4	balance	modal	modal
5	-1	slack, lax	asthenic
6	-2	breathy	breathy
7	open	voiceless	voiceless, aphonia

The terms describe a continuum. Voice terms can be very general, using one term to describe the abducted end and another to describe the adducted end. For instance, breathy or lax may describe an abducted voice and laryngealized or overtight an adducted voice. Voice terms can also be more elaborated, as in Table 2, where two degrees of abduction and adduction are distinguished. For instance, breathy voice is more abducted than asthenic or slack voice and creaky voice is more adducted than tense or strained voice. Even more categories can be distinguished.

The adduction–abduction (or abduction–adduction) continuum may also be described by the terms hyperadduction and hypoadduction (Titze 1994a). These terms tie in with the traditional terms hyperfunctional and hypofunctional. Steps 1–3 in Table 2 are hyperadducted, steps 5–7 hypoadducted. Table 2 merits some further comments:

1 Glottal closure is also known as glottal stop, glottal attack as coup de glotte.

2–3 Creaky voice and its extreme variant creak may be referred to as laryngealized or glottalized voice in phonetic-linguistic description. In phoniatric-logopedic description, especially, creaky voice is referred to as vocal fry or glottal fry. Creaky voice is characterized by the vocal folds being very close to each other, with only a section of them being free to vibrate (Laver 1980, Ladefoged 1988). An extremely adducted voice may sometimes be described as a strangled voice in the phoniatric-logopedic tradition. A mild degree of adduction is characterized as stiff, tense, strained or pressed voice.

4 Modal voice is the difficult-to-define normal voice.

5–6 The term breathy voice is vaguely used sometimes to denote "slack breathiness" (without a noise component) and sometimes "whispery breathiness" (with a noise component, cf. Hammarberg & Gauffin 1994). The distinction between asthenic voice and breathy voice as proposed by Isshiki et al. (1969) avoids the confusion resulting from labelling two types of breathy voice. Asthenic denotes "weakness or lack of power in the voice", whereas breathy "represents a psychoacoustic impression of the extent of the air leakage through the glottis" (Hirano 1981a:83, quoting the Committee for Phonatory Function Tests of the Japan Society of Logopedics and Phoniatrics). However, the use of the term asthenic does not appear to be widespread.

7 An open, voiceless state of the glottis is known as aphonia in the phoniatric-logopedic tradition.

In practice, the term breathy voice can be useful to indicate a type of voice with more abduction (hypoadduction) than in modal voice. A similar cover term for phonation types with more adduction (hyperadduction) than in modal voice can be creaky. Other cover terms have been adopted in other studies (e.g. leaky and strained, Kitzing 1986).

It should be emphasized that the continuum of quality described above does not cover all aspects of voice variation. Variation also results from changes in vocal fold length and stiffness (Sonninen et al. 1992). Nevertheless, the continuum captures an essential part of voice variation. Along similar lines, Klatt and Klatt (1990:820) define voice quality for the purposes of their study as follows:

> "The topic [voice quality] will be restricted to perceptual and acoustic correlates of changes in the breathiness and/or pressed/laryngealized nature of the voicing sound source."

What happens along the breathy–creaky continuum during phonation can be described by several variables (see e.g. Hirano 1988, Gauffin & Sundberg 1989, Sonninen et al. 1992, Sundberg 1994, Titze 1994a, Ladefoged & Maddieson 1996):

1. The intrinsic laryngeal muscles (especially the thyroarytenoid or vocalis muscle) are more contracted in strained voice and even more contracted in creaky voice than in breathy voice.
2. The vocal folds close more rapidly in creaky voice than in breathy voice within each cycle of vibration.
3. In creaky voice the vocal folds are closed longer than in breathy voice within each cycle of vibration.
4. In creaky voice the vocal folds "collide" briskly, whereas in breathy voice they touch softly or may not touch at all (incomplete closure).
5. There is more turbulent air flow through the glottis in breathy voice than in creaky voice.

These phenomena characterize the breathy–creaky continuum. In chapter 3.2 they will be related to spectral properties.

Singing

The discussion above has concentrated on the speaking voice. When singing, the voice is modified in dynamic range, in balance of loudness and in voice quality (cf. Titze 1995), in addition to the obvious changes in the frequency of vocal fold vibration.

A wider dynamic range, both for expressive purposes and for the purposes of adequate sound transmission to the audience, is needed in singing than in speaking. This is accomplished by regulating subglottal pressure and the vibration of the vocal folds on the adduction–abduction continuum. It is also accomplished by regulating the position of the jaw and the lips; with the jaw lowered and the lips protruded a "megaphone effect" (Titze 1995) increases loudness.

In singing, a balance of loudness is required. Sounds differ in inherent loudness. The loudness of each sound has to be balanced by the loudness of adjacent sounds. As sounds produced at different pitches also differ in loudness, the task of balancing loudness is complex.

Singing also requires an aesthetic voice quality. What is regarded as aesthetic obviously depends on musical tradition and context. Voice quality can be regulated by the many settings described above: the phonatory, overall tension, longitudinal, cross-sectional and velopharyngeal settings. In this study the emphasis is on phonatory settings and voice quality.

In the experience of singers and listeners as well as in the research results of voice scientists, certain pitch areas share a similar voice quality. Such areas of similar voice quality are known as vocal registers (see e.g. Large 1973, Hollien 1984, Castellengo, Roubeau & Valette 1985, Keidar, Hurtig & Titze 1987, Titze 1988, Sonninen 1990, Sonninen et al. 1992). Even though there are further registers at the extremes of the vocal range, descriptions of registers usually distinguish two main categories: (1) a low and heavy register; (2) a high and light register. In singing, the low register is often called the chest register and the high register the falsetto. In speaking, the low register is often called the modal register, and the high register the falsetto (as in the singing voice) or upper register. It is generally assumed that the chest register is similar to the modal register: both represent the typical male quality in speech and in low-pitched singing (see e.g. Titze 1994a). The timbre in chest register is characterized as full, resonant and loud, whereas that in the falsetto is characterized as thin, weak and soft (Arnold 1973). The term falsetto is derived from the "false" voice, i.e. a male voice with female qualities (see e.g. Seidner & Wendler 1982).

Such a description betrays a male bias. In a discussion of gender differences in voice, Titze (1994a) laments the scarcity of data on female voices and calls for new data sets and a new orientation toward the adult female voice and children's voice. He even questions the centuries-old position of the adult male voice as a standard (Titze 1994a:xix):

> "All indications are that the adult female voice is a better standard than the adult male voice. The latter appears to be an aberration of the norm [– –]."

In the female voice three registers are usually identified: the chest register, the head register and the falsetto. Sometimes the registers are referred to as the chest, middle (or mixed) and head registers (e.g. Sundberg 1987). Even though different terms are used there seem to be three registers in the female voice: a low, a middle and a high register.

The usage of Titze (1994a), based on the female voice, simplifies the terminology, while helping to appreciate the similarity (and the differences) in female and male voices. He uses the terms chest voice and head voice for both female and male singing. The falsetto is reserved for the female singing voice. Thus, the head voice can be described as a mixture between chest and falsetto. As the male singing voice lacks "real" falsetto, the head voice is often called falsetto.

The registers are bridged by areas of transitions, where an involuntary change from one register to another (also known as the voice break) may occur (Tarneaud 1961, Colton & Hollien 1972, Large 1973, Sonninen et al. 1992). Thus, an ascending pitch series consists of plateaus and (possibly abrupt) transitions. In classical singing abrupt transitions are avoided, striving at the ideal of a voice without registers. On the other hand, yodelling requires abrupt transitions and registers of a very different quality.

These areas of transition are called passaggi in the Italian vocal pedagogy tradition. The transition from chest voice to a higher register (or vice versa) is the first transition (primo passaggio) in female voices and the second transition (secondo passaggio) in male singers. The transition is rather consistently found in the region of 300–350 Hz (D4–F4) in both females and males (Titze 1988). However, the location of the transition depends on several factors: individual variation, voice class (the transition is higher in sopranos and tenors than in altos and basses) and on sex (e.g. 300–350 Hz for males and about 400 Hz (G4) for females). Sundberg (1987) defines the male region of transition as comprising a larger area: 200–350 Hz. The head (middle) register in the female voice borders on the primo passaggio (about 350–400 Hz) and the secondo passaggio (about 660 Hz or E5).

When trying to avoid abrupt transitions singers "cover" their voice (see e.g. Luchsinger & Arnold 1965, Sonninen 1968, Hertegård et al. 1990). On the contrary, nonsingers usually cannot make transitions smooth, thereby revealing the existence of registers with boundaries between them.

Covered singing was introduced as early as the beginning of the 19th century (Luchsinger & Arnold 1965). However, covered is an elusive term. From the point of view of perception, the covered voice has been characterized as dark or sombre (Bloofhooft 1985). From the point of view of production, covering appears to involve a lowering of the larynx and an enlarging of the supraglottal tract (Large 1972). Titze (1994b:337) defines covered voice as follows:

> "A darkened quality obtained by rounding and protruding the lips or by lowering the larynx; the term is likely to stem from covering (fully or partially) the mouth of a brass instrument to obtain a muffled sound; acoustically, all formants are usually lowered and a stronger fundamental is obtained."

Covered voice may share some characteristics with a type of phonation described as optimal for singing: "flow phonation" or "resonant mode"

(Gauffin & Sundberg 1989, Hertegård et al. 1990, Titze 1994a). Flow phonation is characterized by highest possible amplitude of vocal fold vibration and complete closure of the vocal folds in the closed phase of the vibratory cycle. The amplitude is usually higher than in modal voice, except when a person habitually uses flow phonation in speech (Gauffin & Sundberg 1989). In relation to flow phonation, effectiveness decreases both when the glottis is more abducted (i.e. toward breathy voice) and when the vocal folds are more forcefully adducted (i.e. toward creaky voice).

In addition to the term covered, the classically trained singing voice is often referred to as a supported voice (Arnold 1973, Sonninen 1993, Sonninen, Hurme & Sundberg 1994, Titze 1994a, Griffin et al. 1995). However, among singers and voice teachers there is a lack of agreement on the definition of support or breath support and on how it is produced. Many singers report that they feel when they are singing supported and when unsupported. Trained listeners sometimes report that they can hear whether someone is singing with support or without support. Thus, support is a sensation both singers and listeners can recognize. On the other hand, some singers and singing pedagogues resist the concept of support and see it as detrimental to the singing voice.

Even though support is a concept that is not endorsed by all singers and singing pedagogues, it is undeniably used by large populations of singers and voice pedagogues. Physiologically, a highly complex breath control appears the essence of support in singing (Titze 1994a). In other words, support (or lack of support) may be connected with the regulation of subglottal pressure (Gauffin & Sundberg 1989). For singers this may be of little help, as subglottal pressure cannot be monitored by them. The acoustic characteristics of support are not well known (Sonninen et al. 1994, Griffin et al. 1995). The present study aims at describing the spectral properties of supported and unsupported voice.

A further parameter related to the quality of the singing voice is the vertical position of the larynx. In perceptual studies, an undesirable strained or pressed quality of the singing voice has been associated with a raised larynx (Sundberg 1994). Thus, a low larynx position would be advantageous for singers, as the vocal folds would be in a more abducted position. Indeed, several studies on professional singers have concluded that during singing the larynx is in a low position irrespective of pitch (e.g. Shipp & Izdebski 1975, Dmitriev & Kiselev 1979). On the other hand, Johansson, Sundberg and Wilbrand (1985) and Pabst and Sundberg (1992) report that the larynx rises with pitch at least in some subjects, especially at higher pitches. Wang (1983) and Hurme & Sonninen (1995) have noticed much variation in the position of the larynx across singers: some singers raise the larynx with pitch, others do not – and even more complex patterns occur. Such disagreement in the results suggests that while an invariantly low larynx position may be advantageous for singing, it is possible to sing well with the larynx in other positions too. A low larynx position may be more useful in some voice classes than others.

2.5.3 Perception

The perception of voice is probably not as well understood as voice production and acoustics. There are three main perceptual attributes for voice: pitch, loudness and voice quality.

Pitch (or fundamental pitch) is the subjective quality primarily associated with the fundamental frequency of vocal fold vibration. In other words, it is the attribute of auditory sensation in terms of which a sound may be ordered on a scale from low to high. Musical scales are perceptual, with higher and lower notes arranged in various ways in different musical traditions and cultures.

Loudness, in turn, is a perceptual measure of a sound on a strong-weak continuum, associated with the amplitude of vocal fold vibrations. Sound pressure level (SPL, measured in decibels) is a physical correlate of loudness. In music, loudness is indicated by terms like piano and forte, which also serve as instructions to the performer.

Voice quality (or vocal quality or timbre) is not the quality of voice in the sense of the sound resulting from phonation (vibration of the vocal folds). Laver (1980:1) gives the following definition, which emphasizes the perceptual nature of voice quality:

> "Voice quality is conceived here in a broad sense, as the characteristic auditory colouring of an individual speaker's voice, and not in the more narrow sense of quality deriving solely from laryngeal activity. Both laryngeal and supralaryngeal features will be seen as contributing to voice quality. "

There are many voice qualities – and many names or terms to describe them. Voice quality can be described by both specialist terms and non-specialist labels. There are probably hundreds or thousands of metaphorical labels for voices in all languages. Such labels can easily be coined from impressions of voices. A wide and subtle vocabulary can be a "perceptual swamp" (Michel 1985).

From the scientific point of view the use of impressionistic terms is limited, because it is difficult to generalize and understand them outside (and even within) the community in which they have arisen. Scientific terms require much more: there must be an objectively establishable link between the physiological cause, the acoustic consequence and the auditory label that is assigned (see e.g. Sonninen & Hurme 1992). Gross labels would often better be replaced by componential terms: e.g. "hyperfunctional breathy voice with intermittent aphonic moments" describes a voice better than "breathy" or "hoarse" (Laver 1980, Hammarberg & Gauffin 1994). Indeed, hoarse is a very unsatisfactory term, as it fails to give information about vocal fold behavior on the adduction–abduction continuum.

2.5.4 Acoustics

The acoustic manifestation of voice is a product of respiratory, phonatory and articulatory maneuvers. Perception of voice is based on the acoustic signal. The acoustic signal tells about its origin, about the process of production.

The three main acoustic characteristics of voice are fundamental frequency, sound pressure level and spectrum. Fundamental frequency is related to the frequency of vocal fold vibration and sound pressure level to the amplitude of vocal fold vibration. Spectral properties are related both to the glottal source spectrum and to vocal tract transfer characteristics.

The average speaking fundamental frequency (F0) is a common measure of pitch. There is considerable variation in measurements of F0 during speech, partly depending on the instruments used. Fry (1979) reports about 225 Hz for British females and about 120 Hz for British males. In a compilation of 13 studies on F0 in speakers of American English (usually in reading tasks), Baken (1987) reports average F0 values ranging from 189 Hz to 224 Hz for female subjects and from 100 Hz to 146 Hz for male subjects. To take a non-Anglo-Saxon example, Sallinen-Kuparinen (1978) reports mean F0 values in a reading task for 16–17 year old girls and boys attending vocational school to be 211 Hz vs. 137 Hz, respectively, and for those attending senior high school to be 199 Hz vs. 142 Hz, respectively. Such results show that the average F0 varies across and within individuals and cultures.

In speaking the fundamental frequency varies within narrow limits as compared with singing, where it varies from the lowest notes that a bass can produce (35–40 Hz, on the musical scale G0–B0) to the highest notes that a soprano can produce (more than 2000 Hz, about C7; Luchsinger & Arnold 1965). The lowest notes are characterized by a creaky voice and the highest by a whistle-like quality (Luchsinger & Arnold 1965, Titze 1994a).

In studies of voice, the average sound pressure level (SPL) is a commonly used measure of loudness. The average sound pressure level of connected speech lies in the general area of 70 dB (see e.g. Baken 1987), but it is clear that the average SPL depends on factors such as the communicative context, e.g. whether the speaker is addressing a large audience or talking confidentially with one person, as well as on the measurement procedure (e.g. distance of speaker and decibel meter). SPL also differs between singers and nonsingers. Sundberg and Gauffin (1979) reported that the minimum SPL of voices where pitch ranged from 70 to 250 Hz, measured at a distance of 50 cm, was about 75 dB in the piano of a singer, whereas untrained singers were able to phonate at a lower intensity, about 60 dB. The maximum SPLs did not differ much between singers and nonsingers: 93 dB vs. 90 dB. However, the maximum SPLs between singers and nonsingers have also been reported to differ more, up to about 10–12 dB (Titze & Sundberg 1992). The two studies mentioned above show that SPL varies as a function of F0: low pitches are associated with low SPL and high pitches with high SPL.

The main emphasis in this study is on the spectral properties of voice. The radiated spectra are influenced both by vocal tract transfer characteristics and by the glottal source. The spectral properties and thus the identity of the speech sound (or segment) produced are determined partly by the shape of the vocal tract, which is regulated mainly by the tongue but also by other articulators such as the lips, the mandible and the velum. Segmental properties result from short-term adjustments. Spectral properties of voice can also be habitual, resulting from long-term adjustments of the articulators. The acoustic consequences of long-term adjustments of articulators, i.e. voice quality, are complex. They have been described in detail by e.g. Laver (1980) and Nolan (1983). The adjustments include habitual nasalization and a close jaw setting. They also include a lowered or raised larynx and protruded or retracted lips, which lengthen or shorten the vocal tract, thereby altering its resonance properties. For instance, Dmitriev and Kiselev (1979) have shown a connection between the voice class of a singer (e.g. tenor vs. baritone) and vocal tract length, measured by the vertical position of the larynx.

However, the present study concentrates on the effect of the glottal source on spectral properties. The study is indirect, from measurements of long-term averaged spectra of speaking and singing voices. The choice is motivated theoretically (Klatt & Klatt 1990) by the importance of the breathy–creaky continuum for voice variation. It is also motivated by the results of Nolan (1983), Gobl and Ní Chasaide (1992) and Nolan & Kühnert (1995), who find the spectral consequences of phonatory adjustments more distinct than those of articulatory adjustments.

To describe the connection between phonatory settings, spectral properties and the behavior of the vocal folds, the adduction–abduction continuum is recapitulated. Voices leaning toward the breathy end of the continuum are characterized by high rate of air-flow, low vibratory amplitude, slow rate of closing movement and soft collision or incomplete closure of the vocal folds. Voices leaning toward the creaky end of the continuum are characterized by low rate of air-flow, high vibratory amplitude, fast rate of closing movement and brisk collision of the vocal folds. These phonatory phenomena can be linked to the acoustic source spectrum (Fant, Liljencrants & Lin 1985, Gauffin & Sundberg 1989, Kitzing & Åkerlund 1993, Titze 1994a). The glottal waveforms of voices leaning toward the breathy end are characterized long pulse width, low pulse skewing and progressive closure, whereas the waveforms of voices leaning toward the creaky end are characterized by very short pulse width, high pulse skewing and very abrupt closure. Hence, voices toward the breathy end of the continuum are characterized by steep-falling spectral tilt and high turbulent noise. On the other hand, voices toward the creaky end are characterized by flatter spectral tilt and low turbulent noise. (Lee & Childers 1991, cf. Isshiki, Yanagihara & Morimoto 1966.)

Klatt and Klatt (1990:820) identify the following potential acoustic cues to variations in voicing sound source ranging from laryngealized to normal to breathy phonation (i.e. the creaky–breathy continuum): (1) an increase in the amplitude of the fundamental frequency component; (2) an increase in the amount of high-frequency noise that replaces the higher frequency harmonics; (3) an increase in lower formant bandwidths; (4) the introduction of extra poles and zeros in the vocal tract transfer function (due to tracheal coupling).

The relative prominence of the voice fundamental has been observed in a number of studies on breathy voice: breathiness is correlated with a relatively high level of the fundamental frequency in relation to the level of the first formant or the second harmonic (e.g. Fischer-Jørgensen 1967, Bickley 1982, Huffman 1985, Ladefoged & Antoñanzas-Barroso 1985, Hammarberg, Fritzell, Gauffin & Sundberg 1986, Klatt 1987, Klatt & Klatt 1990, Kasuya & Ando 1991, De Krom 1994 and Trittin & de Santos y Lleó 1995). Holmberg, Hillman and Perkell (1995:179) maintain that "the degree of glottal adduction was reflected relatively well in AH1–AH2" (their abbreviations for the levels of the fundamental and the second harmonic).

Breathy voices are also associated with high-frequency noise, due to the persistence of an incomplete closure of the vocal folds during the closed phase (e.g. Klatt 1987, Dejonckere & Wieneke 1992, Fant 1993). An incomplete closure of the vocal folds is both a female and male characteristic. However, it is more common in the female voice (Bless, Biever & Shaik 1986, Södersten 1994).

As Hammarberg (1992) points out, there are two types of breathy voice: one which is breathy and hypofunctional (with a dominant fundamental frequency in the spectrum) and another which is breathy and hyperfunctional (with noise or turbulence in the upper spectrum). If the two need to be differentiated, the former can be called breathiness and the latter whisperiness (Hammarberg et al. 1986). Isshiki et al. (1969) probably intend the terms asthenic and breathy to capture these two types of breathy voice.

In all, a number of spectral correlates have been demonstrated for the breathy–creaky continuum. They will be discussed in more detail in chapter 3.2, when the spectral correlates of various types of voices are examined.

The diversity of the spectral cues for the breathy–creaky continuum has a parallel in the cues for nasality. The perception of nasality depends on highly complex and variable acoustic cues. These include extra poles and zeros due to nasal coupling and an increased bandwidth of formants (e.g. Hattori, Yamamoto & Fujimura 1958) – cues shared by breathy voices. Nasal voices are also characterized by a lower overall amplitude (Bernthal & Beukelman 1977, Kyttä & Hurme 1982). A low overall amplitude is a characteristic of nasal voices, whereas breathy voices are characterized by the fundamental dominating the spectrum (resulting in a steep spectral slope).

Even though much acoustic research has been carried out on the relationship between voice production and voice acoustics, much remains to be done. The present study is a contribution to the description of the spectral properties of various voices. The method employed is long-term average spectrum analysis (LTAS).

3 VOICE ACOUSTICS

3.1 Long-term spectrum analysis

The methods developed for the study of speech, i.e. phonetic and phoniatric-logopedic methods, can also be used for the study of voice. The methods used in voice research can be divided into physiological, perceptual and acoustic, investigating respectively the production, perception and acoustics of voice.

There are many physiological methods available for the study of phonation and articulation. The vocal folds can be examined with the aid of e.g. ultra-high speed cinematography, fiberoptic endoscopy, stroboscopy, photoglottography, electroglottography, electromyography and ultrasound (see e.g. Hirano 1981a, Baken 1987), as well as by radiographic methods (Sonninen 1956, Sonninen et al. 1992). The vertical movements of the larynx can be registered by a twin-channel electroglottograph (Pabst & Sundberg 1992, Rothenberg 1992, Sonninen et al. 1994). In addition to the vertical position, the anterior-posterior or sagittal movement of the larynx can be measured by means of radiography (Hurme & Sonninen 1995).

The perception of voice is often studied in listening tests (see e.g. Wendahl 1963, Scherer 1982, Anders, Hollien, Hurme, Sonninen & Wendler 1988, Klatt & Klatt 1990). Listening tests undoubtedly give information about the stimuli. However, the results of listening tests need to be interpreted with caution: even professional raters show much variation in their judgements (Kreiman, Gerratt & Precoda 1990, Minifie, Green, Smith & Huang 1995). Besides, peer-perception of voices may differ considerably from professional judgements (see e.g. Lass, Ruscello, Stout & Hoffman 1991). Such results emphasize the need for objective studies of the voice by means of physiological and acoustic methods.

Acoustic methods

As they are noninvasive and cause minimal disturbance to the speaker or singer, acoustic methods permit the examination of voice produced in relatively natural circumstances, while yielding objective, quantitative data about voice characteristics. In acoustic studies of voice, the signal radiated from the lip opening is usually investigated. However, there are also methods for extracting the unfiltered glottal waveform. Two well-known examples are the Sondhi tube or reflectionless tube (Sondhi 1975, Monsen & Engebretson 1977) and inverse filtering (Miller 1959, Rothenberg 1973, Hertegård et al. 1990, Hertegård 1994, Alku & Vilkman 1995). Such methods aim at eliminating the effect of the resonator, thus arriving at the glottal waveform. The information obtained describes the voice source, which is valuable for many research purposes.

Four basic acoustic parameters are often extracted from a speech signal: duration, amplitude, fundamental frequency and spectrum. It is possible to extract several further acoustic parameters, often by combining basic parameters. For instance, measures of amplitude perturbation and pitch perturbation result from combining measures of duration to measures of sound pressure level and fundamental frequency (see e.g. Lieberman 1963, Askenfelt & Hammarberg 1986, Hirano 1989). There are also methods for describing the noise or turbulence in unvoiced speech signals or in speech signals consisting of both voiced and unvoiced signals (see e.g. Kasuya, Ogawa & Kikuchi 1986, Baken 1987). Such further parameters are useful in studies of e.g. dysphonic voices.

The sound pressure level (SPL) of signals is measured using a sound level meter (decibel meter) or a level recorder. SPL cannot be measured from a tape recording, unless it has been registered at the time of original voice production. Hammarberg (1985) describes a system, implemented routinely in clinical work, which utilizes analogue technology to make audio recordings with information on SPL. With digital recordings (DAT), information on sound pressure level can be obtained with ease (see e.g. Laukkanen, Vilkman, Alku & Oksanen 1996).

The fundamental frequency of a voiced signal can be measured by many techniques, both analogue and digital (see e.g. Hess 1983, Baken 1987). While it is easy for the ear to hear pitch changes in a voiced signal, the task of detecting F0 may be difficult for a machine, especially when the signal contains unvoiced components (as is often the case in voice disorders), when fundamental frequency varies rapidly (as often in speech) and when the range of F0 variation is large (as often in singing). In fact, both dedicated analyzers and computer programs frequently make mistakes in F0 detection.

Sound pressure level measurements of sustained phonation can be made when a person produces maximum and minimum vocal intensity at selected pitches. The result is a voice range profile, also known as a phonetogram (Damsté 1970, Coleman, Mabis & Hinson 1977). Attempts have also been made to describe the variation in SPL in connected speech by registering cycle-by-cycle variation in SPL across fundamental frequency (Sonninen, Hurme & Toivonen 1988). Phonetograms have been complemented by other measures, such as signal-to-noise ratio (Klingholtz 1990) and spectral information (Gramming & Sundberg 1988, Pabon & Plomp 1988).

These acoustic measures of voice serve many research purposes. The present study, however, relies mainly on spectrum analysis.

Spectrum analysis

A spectrum is a Fourier transform of a sound signal, showing the amplitude of the harmonics in the decibel scale as a function of frequency, usually in Hz. Thus, vocal tract resonances or formants are revealed. A number of techniques are available for spectrum analysis. These include variants of sound spectrography: analogue and digital, as well as implemented in computer programs. It is possible to compare the spectra of e.g. sustained vowels under various conditions, such as when singing and speaking. Spectrum analysis can also be carried out on longer signals by taking a number of samples of the signal and calculating the mean of all the accumulated spectra.

In addition to examining the formant structure of sounds or longer stretches of speech, spectral analysis may also aim at indirectly examining the properties of the voice source. For that purpose, various measures of the spectral slope are used. The spectral slope can be measured both from radiated voice (e.g. Löfqvist & Mandersson 1987, Ladefoged, Maddieson & Jackson 1988) and from the source spectrum by glottal inverse filtering (e.g. Bickley 1982, Javkin & Maddieson 1982, Gobl & Ní Chasaide 1992, Titze & Sundberg 1992).

Spectrum analysis provides an indirect window on both laryngeal activity and on articulatory movements. While spectra represent the joint contribution of the sound source and the vocal tract transfer function, suitable measurements applied to spectra can provide information both on the vocal tract and on the voice source (see e.g. Löfqvist & Mandersson 1987). Such measurements are based on standardization of the contribution of e.g. the vocal tract by producing the same sound with variations in the voice source. The contribution of the vocal tract can also be standardized instrumentally by computing the average spectrum of a longer stretch of speech.

Long-term spectra

Long-term average spectrum (LTAS) analysis, also known as long-term spectrum (LTS) analysis, describes the long-term characteristics of voice or speech by calculating the average of all spectra from a speech sample. More technically, a summation spectrum is calculated, as a true long-term average spectrum cannot be calculated by Fast Fourier Transforms for signals longer than about 0.1 seconds (Klingholtz 1990).

The idea of investigating speech and voice by means of average long-term spectra goes back a long way. The first LTAS studies were probably carried out by Crandall and MacKenzie (1922), Sivian (1930) and Dunn and White (1940, cf. Kosiel 1973). Possibly due to the difficulty and cost of constructing long-term spectrum analyzers, the method did not become popular until the 1960s and 1970s (Carr & Trill 1964, Li, Hughes & House

1969, Blomberg & Elenius 1970, Kosiel 1973 and Ostwald 1973). In the 1970s commercial real-time spectrum analyzers with averaging capabilities became available. The method was applied to several areas of voice research, such as dysphonia (e.g. Fritzell, Hallén & Sundberg 1974, Prytz & Frøkjær-Jensen 1976, Hammarberg et al. 1980, Hurme 1980b), effects of voice training (e.g. Wedin, Leandersson & Wedin 1978), and comparison of languages (e.g. Majewski et al. 1977).

The instruments for long-term spectrum analysis have developed over the years from laboratory-specific analyzers to commercial analyzers (e.g. the Brüel & Kjær 2031 Narrow Band Spectrum Analyzer used in the present study). Computer control and post-processing can be added to commercial analyzers (see e.g. Hurme & Pirinen 1984, Harmegnies & Landercy 1985, Hurme & Sonninen 1985b). Digital computers can be programmed to perform long-term spectrum analysis as well as a wide variety of other analyses (e.g. the Signalyze program used in the present study; Keller 1992 & 1994).

Long-term spectrum analysis calls for the elimination of pauses and unvoiced segments in the signal prior to analysis. Otherwise the average amplitude levels of connected speech are quite different from those of sustained phonations or isolated monosyllables, as connected speech contains both pauses and unvoiced segments (see e.g. Bricker 1965, Hurme 1980b).

The long-term spectrum is averaged from a sample of continuous speech. It has been proposed that 10 seconds is enough to reach a stable spectrum (Furui, Itakura & Saito 1972). Usually, however, time periods of 30 to 60 seconds have been proposed (e.g. Li et al. 1969, Fritzell et al. 1974). Hurme and Pirinen (1984) also suggest that the effect of individual speech sounds is no longer significant after about 30 seconds of continuous speech.

Advantages and disadvantages

As any data collection method, the long-term average spectrum analysis has advantages and disadvantages (Kitzing 1986, Löfqvist 1986, Löfqvist & Mandersson 1987, Klingholtz 1990, Kitzing & Åkerlund 1993). The advantages include ease of data collection: short and long voice samples can be equally analyzed. During the analysis of long samples, the effect of individual sounds disappears. LTA spectrum analysis is noninvasive. LTA spectra give a fairly simple description of a phenomenon which is very complex.

The disadvantages include difficulty of interpretation and difficulty of measurement. To interpret LTA spectra, an understanding of voice acoustics is essential. Even though voiceless sounds are typically eliminated from LTA spectrum analysis, voiced sounds may contain voiceless components which contribute to the LTA spectrum. The contribution of voiced and voiceless components cannot be seen in LTA spectra. Neither are there any standardized measurements that are generally applied to LTA spectra. As described in this study, many measures are available. The choice among them depends on theoretical considerations and on the research purpose. Long-term spectrum studies have also been criticized for not taking into account vocal intensity (Kitzing 1986, cf. Sonninen & Hurme 1982). The spectra of loud and soft voice, for example, differ substantially (Fant 1959).

Long-term average spectra (LTAS) can give information on the voice source (Löfqvist & Mandersson 1987:221):

> "The long-term average spectrum provides information on the spectral distribution of the speech signal over a period of time. Such spectra have been used for studies of the human voice source. The speech signal represents the product of sound source and the vocal tract transfer function. The latter differs for different sound segments, but in the averaging process the short-term variations due to phonetic structure will be averaged out and the resulting spectrum can be used to obtain information on the sound source; if the analysis is restricted to voiced sounds, the sound source is the vibrating glottis. In order to further minimise variations due to phonetic structure, the analysis can be made of the reading of a standard text. Thus, while the long-term spectrum represents the joint contribution of the sound source and the vocal tract transfer function, suitable measurements applied to the spectrum can provide information on the voice source."

Long-term spectrum analysis describes indirectly both laryngeal and supralaryngeal behavior, as shown in studies where deliberate articulatory and phonatory settings were assumed and their effect on long-term spectra investigated (e.g. Nolan 1983, Harmegnies, Esling & Delplancq 1989, Nolan & Kühnert 1995). LTA spectra are more sensitive to phonatory than articulatory settings (Nolan 1983:152): " [– –] in general gross long-term spectral changes do not result from supralaryngeal settings."

The effect of supraglottal factors can be minimized by standardizing the material to be analyzed. For instance, all subjects can read the same text or phonate the same vowel. Such restrictions increase the possibility of extracting useful information on phonatory behavior.

It is evident that LTAS is not a universal tool that can be applied for any purpose. Its use requires caution. It is valuable in intraindividual comparisons, as interindividual variation between LTA spectra is substantial (Hurme & Pirinen 1984, Kitzing 1986). LTAS analysis is especially suitable for comparing repeated measurements for one subject. It can be used to monitor the effect of voice therapy (e.g. Södersten & Hammarberg 1993). It has also been used in studies on vocal fatigue (e.g. Ohlsson, Järvholm & Löfqvist 1987).

Long-term spectrum measures

Long-term spectra can be evaluated qualitatively (e.g. Nolan 1983), looking for similarities and differences and by grouping the spectra accordingly. However, a quantitative approach is more common, as it makes statistical treatment and testing possible. Many quantitative measures or sets of measurements have been proposed to describe long-term spectra.

Prytz and Frøkjær-Jensen (1976) proposed the ratio of energy below and above 1 kHz (the alpha parameter) to be a measure of the spectral slope. Nolan (1983) experimented with several measures and decided to use the ratio of energy below and above 1.5 kHz as a measure of slope. However, it would seem that these measures have not been widely used in LTAS studies (Pittam 1994).

To study normal voice and singing voice, Gauffin and Sundberg (1980, 1989) introduced a set of measurements of sound levels in four spectral bands; the lowest included the fundamental and the second, third and fourth were 0.4–1, 1–2 and 2–4 kHz, respectively. Hammarberg and her co-workers developed two related sets of measurements for the analysis of dysphonic voices. The first (Hammarberg, Fritzell & Schiratzki 1984) consists of 5 measures: the peak level between 400 and 600 Hz, normally corresponding to the averaged overall sound pressure level of the voice, and the difference between this dB level and the following four levels: the fundamental, 1.5 kHz, 5 kHz and the highest point between 5–10 kHz. Another set of measurements (e.g. Hammarberg et al. 1980) consists of the maximum level of intensity in three bands, 0–2 kHz (low), 2–5 kHz (middle) and 5–8 kHz (high), complemented by the level of the fundamental frequency. Other measures for describing LTA spectra have been proposed by e.g. Hurme (1980a), Wendler, Doherty and Hollien (1980) and Löfqvist (1986).

Long-term average spectrum analysis is a method that has been and is used frequently. However, there is no generally accepted way to measure the spectra. The spectral measures need to be adjusted to the particular goal of each study.

3.2 Spectral correlates of voice

This study examines dysphonic voices, voices varying in vocal intensity and the singing voice by means of long-term spectrum analysis. In addition, spectral differences between female and male voices are examined. This chapter presents an overview of the spectral correlates of dysphonic voices, vocal intensity and the singing voice as well as female and male voice.

3.2.1 Dysphonic voice

A person's voice shows both invariance, more or less permanent characteristics typical of the person, and variance, more short-term variation of voice in different situations and at different times. The borderline is elusive: voices are subject to change as a result of environmental and social factors and/or as a result of training. A teacher's voice may get tired during the working day as he or she speaks in a loud voice in a noisy environment. As a result of habitual abuse, the vocal folds may develop nodules or other disturbances and voice quality will change. Disordered or dysphonic voices differ from healthy voices in that they disturb the speaker or singer. (Sometimes they also disturb the listener, but this does not necessarily mean that they disturb the speaker or singer.)

There are many kinds of voice disorders and there are many reasons for voice disorders (see e.g. Luchsinger & Arnold 1965, Stemple 1984, Sies 1987, Sataloff 1991). Some disorders are articulatory (e.g. excessive nasality), some phonatory. There are many acoustic studies comparing groups of dysphonic speakers. Such groups may have been formed on the basis of phoniatric diagnosis and/or perceptual evaluation. For instance, Hirano, Tanaka, Fujita and Terasawa (1991) report a number of differences in several fundamental frequency and sound pressure level measurements in a large sample of dysphonic speakers classified according to diagnosis. Many studies reveal that disordered voices often show more irregular variation in the speech waveform than normal voices (e.g. Lieberman 1963, Davis 1979, Murry & Doherty 1980, Laver et al. 1986). Such frequency and amplitude perturbation is known as jitter and shimmer.

Spectral differences have also been observed. Early studies often showed more high-frequency turbulent noise in disordered voices than in normal voices (e.g. Isshiki et al. 1966, Yanagihara 1967). However, such early work often pooled all kinds of dysphonic voices into hoarse voices, failing to differentiate various types of dysphonia.

Deliberately oversimplifying, dysphonias can be divided into two types, those with too much abduction and those with too much adduction. More abduction decreases resistance at the vocal folds, whereby the closed phase in vocal fold vibration becomes shorter (Lindqvist Gauffin 1972, Gauffin & Sundberg 1989); such hypoadduction may also lead to an incomplete closure of the vocal folds. The spectral consequences of increased abduction include a prominence of fundamental frequency and a steep spectral slope. A possible cause for incomplete closure is a partial or complete paralysis of the recurrent nerve. The muscles adducting the vocal folds are not innervated or are innervated only partially. The result is an incomplete closure of the vocal folds: the vocal folds vibrate but do not close fully in the "closed" period. Turbulent noise may be created at the glottis when the closure is incomplete. So little resistance may be offered that the vocal folds do not vibrate at all, with detrimental effects on voice production. Also an inflammation in the vocal folds may result in hypoadduction.

More adduction or hyperadduction tends to increase resistance in the vocal folds and increase the closed phase in vocal fold vibration. Spectral consequences include a shallow spectrum (i.e. less steep than in hypoadduction) and a weak level of the fundamental. Extreme adduction may result in irregular vibrations or even in a suppression of vibrations. Vocal fold resistance may increase for many reasons: inflammation, vocal nodules and laryngeal cancer, among others. The connection between diagnosis and voice characteristics is not simple: e.g. an inflammation in the vocal folds may result in both hypoadduction and hyperadduction.

Spectral differences between various groups of dysphonic speakers or between voices simulating voice disorders have been investigated by many scientists (e.g. Yanagihara 1967, Gauffin & Sundberg 1977, Hammarberg et al. 1980, Wendler et al. 1980, Sonninen & Hurme 1984, Hurme & Sonninen 1985a, 1985b, 1986, Hammarberg 1986, Kitzing 1986, Löfqvist & Schalén 1991, Hammarberg & Gauffin 1994). Such studies have described a number of spectral and other acoustic differences between various groups. In general, a breathy voice is characterized by a steep spectral slope and a strained voice by a shallow spectral slope. For instance, Gauffin and Sundberg (1977) have shown a significant correlation between a measure describing the overall spectral slope (level difference of 0–2 kHz and 2–5 kHz regions) and perceptual ratings on a breathy–overtight (i.e. breathy–creaky) continuum. In relation to the 0–2 kHz level, the level of the 2–5 kHz band is lower in breathy voice than in overtight voice. Hammarberg et al. (1980) report that breathy and hypofunctional voices typically have a low level in the spectrum at the 2–5 kHz region compared with the 0–2 kHz region. On the other hand, in hyperfunctional and overtight voices the difference in the spectral levels at 2–5 kHz and 0–2 kHz is smaller than in breathy and hypofunctional voices. The spectral level at the 5–8 kHz region is high in whispery breathy voices, low in lax or asthenic breathy voices. Hammarberg et al. (1986) report that the level of the fundamental is high in hypofunctional and lax voice and low in hyperfunctional and tense voice. Löfqvist and Schalén (1991) have observed two spectral consequences of acute laryngitis as compared with normal voice: more high-frequency (5–8 kHz) energy and a less steep spectral tilt (measured by the ratio of energy below and above 1 kHz).

Kitzing (1986) has investigated the ability of a large number of acoustic measures derived from long-term average spectrum analysis to differentiate simulated leaky, strained and soft voices from normal sonorous voices. The most potent criteria were: (1) several measures of spectral slope (e.g. 0–1/1–5 kHz and 0.3–0.8/1.5–2.0 kHz ratios) and (2) the difference in level of the fundamental frequency and the first formant region. Leaky and soft voices had on the average a steeper spectrum and a higher level of the fundamental than normal voices. Strained voices had a more shallow spectrum than normal voices.

Thus, there are several potential candidates for the spectral correlates of dysphonic voices. They include measures of the spectral slope in a low region and in a high region, as well as measures of differences in level in low harmonics. The present study compares the spectral correlates of various Finnish dysphonic voices by means of long-term spectrum analysis, looking for similarities and differences by utilizing a wide range of measures.

Spectral studies and other acoustic studies of dysphonia are motivated by the prospect of constructing devices with which to detect dysphonias. Such devices would be useful in the screening of large populations (e.g. Laver et al. 1986). Another useful device would give feedback about a person's voice, for instance in connection with voice therapy. Such endeavours need to be built on solid observations of voice acoustics, including spectral properties.

3.2.2 Vocal intensity

In addition to voice disorders, this study deals with the spectral consequences of vocal intensity. Variation in vocal intensity can be described perceptually, physiologically and acoustically. Voices varying in loudness can be perceived as e.g. soft, normal and loud. These characterizations were used when collecting the material investigated in the study. Physiologically, voices vary in e.g. the amplitude of vocal fold vibration (e.g. Holmberg, Hillman & Perkell 1988, Schulman 1989) and subglottal pressure (Sundberg, Titze & Scherer 1993). Acoustically, voices vary in overall amplitude or sound pressure level (SPL), measured in decibels (dB). When providing the material for the study on vocal intensity, the speakers were instructed to try to speak at prescribed sound pressure levels by monitoring themselves using a decibel meter.

In a study by Coleman et al. (1977), where female and male subjects sustained a vowel at intervals of 10% of their pitch range at minimum and maximum sound pressure levels, the maximum differences between single productions were more than 70 dB at a microphone distance of 15 cm. Mean maximum SPL was 113 dB in women and 117 dB in men, and mean minimum SPL 55 dB in women and 58 dB in men. The sound pressure level of connected speech lies in the range of 70 dB (measured at 15–30 cm, Baken 1987).

Loud and soft voices do not differ only in sound pressure level. They also differ in fundamental frequency and spectral properties. In general, the louder the voice, the higher the fundamental frequency (Coleman et al. 1977, Gramming, Sundberg, Ternström, Leanderson & Perkins 1988, Glaze, Bless & Susser 1990). Another observation, which is more important to the present study is that the louder the voice, the greater the proportion of higher harmonics. In a classical study, Fant (1959, see also 1973) investigated the connection between sound pressure level and spectrum. Long-term spectra were accumulated of the first five Swedish digits and of a word containing the vowel [a]. The speech samples were spoken in three loudness conditions: normal, -10 dB and +10 dB. The results show that what acoustically differentiates the loudness levels is the slope of the spectrum. In soft voice the slope is steep, dominated by the fundamental frequency. In loud voice, the slope is less steep; the harmonics above the fundamental may have an equal or even a higher level than the fundamental. Fant (1959) also observed that the harmonics at about 2 kHz (in digits) or about 3 kHz (in vowels) are more pronounced in loud voice.

Similar results have been obtained by other researchers, e.g. Bricker (1965), Brandt, Ruder and Shipp (1969) and Gauffin and Sundberg (1989). Södersten, Lindestad and Hammarberg (1991) obtained a systematic difference in level between the fundamental and the first formant: the louder the voice, the higher the level of F1. (There was also a clear difference between female and male subjects.)

In a comparison of the spectral consequences of varying vocal intensity (Gauffin & Sundberg 1989), a singer and a nonsinger showed on the whole a pattern similar to that presented by Fant (1959). However, when singing in a loud voice the singer has a higher level in the frequency range of 2–4 kHz, probably due to the singer's formant (Sundberg 1974). In a study of singers who varied their vocal intensity Sundberg (1973) observed that the major effect on the spectrum was in the low frequency part of the spectrum, below 1 kHz. The spectral slope was steeper below 1 kHz in piano singing than in forte singing.

Thus, the configuration of the spectrum is highly dependent on vocal intensity. The present study investigates the spectral consequences of vocal intensity in a Finnish material. This study also tests the LTAS method by systematically studying the influence of variation in voice intensity on the configuration of the long-term spectrum, as suggested by Kitzing and Åkerlund (1993).

3.2.3 Voice types in singing

The spectral correlates of the singing voice have been sought by comparing it to the speaking voice. Acoustically, the singing voice – like the speaking voice – is characterized by many variables, including fundamental frequency, overall sound pressure level, amplitude of harmonics and frequency of formants as well as frequency and amplitude modulation (see e.g. Bloothoft 1985, Sonninen 1987, Sundberg 1987, Rossing 1990, Sundberg, Gramming & Lovetri 1993). A prominent characteristic of some singers in comparison with nonsingers is the singer's formant around 3 kHz (Sundberg 1974, Seidner & Wendler 1982, Estill, Baer, Honda & Harris 1985, Sundberg 1987, Rossing 1990).

As described above on page 33, both maximum and minimum overall sound pressure levels are higher in singers than nonsingers. The observed differences in SPL between singers and nonsingers can be explained as follows (Titze & Sundberg 1992). When raising pitch, vocal folds are elongated. Elongated vocal folds require a higher driving pressure than shorter and more lax vocal folds. For that purpose, a higher subglottal pressure is needed. As subglottal pressure mainly determines SPL, the result is higher SPL in higher pitches (see e.g. Sundberg 1995). However, singers and nonsingers appear to differ in the way they sing at higher pitches: singers presumably "learn to lower their effective glottal impedance to transfer more power from the source to the vocal tract for a given lung pressure" (Titze & Sundberg 1992:2946). Titze and Sundberg discuss two mechanisms that singers can use, the first phonatory and the second articulatory. When phonating, singers can use a different mode of vibration that produces greater amplitudes of vibration for the same driving pressures (i.e. less energy lost in the tissue due to less adduction in the vocal folds). They can also make an articulatory adjustment in the lower pharynx, near the false vocal folds, that produces a more favourable input impedance of the vocal tract. Nonsingers, on the other hand, cannot make these adjustments; therefore, they cannot sing in as loud a voice as singers and their voices tend to become more strained or tense (i.e. too adducted) at higher pitches.

The spectral correlates of the singing voice can also be sought by comparing various kinds of singing voices to each other. Perceptual characterizations of voice have been associated with spectral properties in many studies (e.g. Luchsinger & Arnold 1965, Bloothooft 1985). An early example is provided by Talvi (1931:50):

> "Såsom hörande till ansatsrörets akustik böra vi också i korthet nämna de akustiska orsakerna till de viktigaste s.k. röstfärgerna. Mina akustiska undersökningar giva vid handen, att: "Ljus" kallas en ton med starka, höga övertoner [– –]. "Mörk" är en ton, som antingen saknar övertoner eller endast har de första övertonen tydliga. "Mjuk" är en ton, där den första övertonen [– –] dominerar och de andra antingen saknas eller äro mycket svaga. "Glansfull" är tonen, när den tredje övertonen [– –] dominerar."

To take another example, Järvelä (1991) reports that "cold" voices have more amplitude in the higher harmonics in comparison with "warm" voices. The results of such studies vary considerably, as it is difficult to agree on the meaning of the labels used to describe voices.

Female and male singers differ in spectral properties. The singer's formant is typically a male characteristic, even though low female voices may also have such a formant (Seidner, Schutte, Wendler & Rauhut 1985, Bloothoft 1985). Another difference between female and male singers is the dominance of the fundamental in the female voice (White 1989). In a comparison of alto and tenor singers, a dominant fundamental was observed in the alto voice and a relatively weak fundamental in the tenor voice (Ågren & Sundberg 1978).

The spectral correlates of vocal registers in singing have also been examined. For instance, Keidar et al. (1987, see also Colton 1972) observed a correlation between spectral slope and perceived register. The voice samples identified as chest voice had a shallower spectral slope; those identified as falsetto had a steeper slope.

There are many traditions of describing the quality of the singing voice. For instance, Estill et al. (1985) examined the spectra of voice qualities characterized as speech, twang, falsetto, opera, low larynx and belting. Such qualities include many varieties of singing. However, there are two main traditions of describing the quality of the classical singing voice. One characterizes good singing as covered, poor singing as open (see e.g. Luchsinger & Arnold 1965, Sonninen 1968, Hertegård et al. 1990, Miller & Schutte 1994). The other tradition describes good singing as supported, poor singing as unsupported (see e.g. Sonninen 1993, Sonninen et al. 1994, Griffin et al. 1995).

The spectral correlates of covered and open voice received attention as early as 1912. Pielke (quoted by Luchsinger and Arnold 1965) associated a strong fundamental, a weak second harmonic and generally rich harmonics with covered singing and a prominent second harmonic with open singing. Luchsinger (1951:67) reported acoustic observations on covered and open singing: in covered voice, 9 out of 14 subjects showed an increased level of the fundamental and more overtones. However, his examples suggest that sound pressure level is higher in covered voice than in open voice; even though there may be more harmonics in the spectra of covered voice, in relation to the level of the fundamental they appear to be lower in level in covered voice than in open voice.

In a study focusing on length-changes in the vocal folds, Sonninen (1968) reported results on acoustic measurements of covered and open voice. However, his report is not explicit regarding spectral slope, though one can infer from the narrow-band spectra reproduced in the article that the vowels sung in open voice had stronger harmonics than those sung in covered voice.

Levels of harmonics and spectral slope can be regulated in the larynx on the adduction–abduction continuum. Levels of harmonics can also be regulated by the vocal tract by shifting formants lower and higher (see e.g. Sundberg 1987, Bloothoft & Plomp 1988, Miller & Schutte 1994). For instance, by singing a low vowel more [æ]-like the second formant is shifted up in frequency. If a low vowel is sung more [ɔ]-like, the two lowest formants are shifted down in frequency. Such shifts raise and lower the levels of harmonics, and thus change the perceptual impression.

Differences in the levels of the harmonics between covered and open voice may be due to glottal behavior and/or to articulatory maneuvers. A covered voice can be less adducted than an open voice or a singer can adjust or "tune" the resonance characteristics of the vocal tract to increase or decrease the amplitude of a certain harmonic (Miller & Schutte 1994). Hertegård et al. (1990:226) have estimated the relative contributions of these factors. They estimate that less than one third of the difference of 5 dB in the level of the fundamental between covered and open voice observed in their study can be explained by the about 20% lower first formant frequency in covered singing; in their interpretation the rest of the difference is probably due to changes in the voice source.

Hertegård et al. (1990) have studied covered and open voice by means of fiberoptics, inverse filtering and acoustic analysis. They observed differences in formant frequencies of the vowel [æ], sung near the register transition area (275–400 Hz in their 11 subjects). The two lowest formants had higher frequency in open voice than in covered voice. On the other hand, there was no difference in the frequency of the singer's formant between covered and open voice. They also observed differences in the levels of harmonics. The fourth harmonic was higher in covered voice than in open voice – a result of the tuning of the second formant to the fourth harmonic, in their interpretation. The amplitude of the fundamental frequency (L1) was higher in covered voice than in open voice. On the other hand, the amplitude of the second harmonic (indicated by L2 in the present study) was lower in covered voice than in open voice. Consequently, in relation to the level of the fundamental frequency L2 was lower in covered voice and higher in open voice.

The level difference between the first and second harmonics appears to be a spectral correlate of covered vs. open voice. The role of the higher harmonics is unclear. Articulatory adjustments may contribute to the covered vs. open distinction. However, a difficulty in comparing covered and open voice is that while covered voice can be equated with a good, trained voice, open voice may mean very different types of voice in different studies.

Another distinction often made when describing the singing voice is supported vs. unsupported. The acoustic characteristics of supported and unsupported voice have been studied by Griffin et al. (1995) in connection with physiological measurements. In all subjects, supported voices had a higher sound pressure level than unsupported voices. This is a result expected on the basis of earlier observations according to which voices with support have a higher subglottal pressure than voices without support (Gauffin & Sundberg 1989). Griffin et al. (1995) also observed differences in the fourth formant: in all subjects, the frequency of the fourth formant was lower in supported voice than in unsupported voice. In male subjects, the amplitude of the fourth formant was higher in supported voice than in unsupported voice. These results are probably connected with the gender-specificity of the singer's formant. Sonninen et al. (1994) reported preliminary observations of the level difference between the first formant and the fundamental: in relation to L1 the amplitude of the first formant was systematically higher in supported voice than in unsupported voice. This pattern was clearer in male singers than in female singers.

The spectral correlates of supported vs. unsupported voice remain unclear. The results of Sonninen et al. (1994) suggest the low frequency range as a candidate for distinguishing supported from unsupported voice.

In conclusion, acoustic studies of such distinctions as covered vs. open and supported vs. unsupported are far from conclusive; the results can even be controversial. The difficulty of defining these distinctions both physiologically, acoustically and perceptually may lead to heterogeneous groups (and to much variation in acoustic measurements). Nevertheless, various measures of spectral slope appear useful in the study of the singing voice. Additional measures, such as formant frequencies are also useful. In fact, Miller and Schutte (1994) see a covered transition from chest to head voice primarily as an articulatory phenomenon, produced by positioning the first formant appropriately, and not as a phonatory phenomenon, produced by changing the balance of intrinsic muscles in the larynx. Singers probably use any available method to equalize registers, both laryngeal and articulatory, in order to be able to sing well.

The present study examines the spectral correlates of both covered vs. open and supported vs. unsupported voices in the context of register transition. Finnish singers have produced the samples of good (covered, supported) and poor (open, unsupported) singing.

3.2.4 Female and male voice

Three of the four materials investigated in this study comprise both female and male voices. (The exception is the material on covered vs. open voice, produced by one female singer.) Thus, there is material for comparing female and male voice.

Female and male voices have been described in terms of fundamental frequency and formant frequencies in a number of languages (see e.g. Petersen & Barney 1952, Fant 1959, 1975, Fischer-Jørgensen 1967, Nordström 1977). Systematic differences in both fundamental frequency and formant frequencies have been observed.

In overall amplitude the difference between women and men has been characterized as small (Coleman et al. 1977). On the other hand, Fletcher (1972) has reported women to use lower sound pressure levels in connected speech.

Even though the basic differences between female and male voices are fairly well known in terms of fundamental frequency and formant structure, a description of that kind is not adequate. For instance, attempts at synthesizing female voices on the basis of the results of such analyses have yielded poor results (e.g. Karlsson 1992). Female and male voices apparently differ in some additional dimensions.

One such dimension is the adduction–abduction continuum, manifested in a steeper spectral slope and more breathiness in female voices than in male voices. An anatomical explanation for the differences in spectral slope between women and men has been given (e.g. Monsen & Engebretson 1977; see also Vilkman et al. (1995) for a more complex "critical mass" theory). On the average, men have longer vocal folds and more mass in the vocal folds, whereas women have shorter vocal folds with less mass. It is conceivable that longer and more massive vocal folds induce a stronger collision of the vocal folds and the shorter and less massive vocal folds a weaker collision (or even no collision, i.e. incomplete closure of the vocal folds). A strong collision leads to more amplitude in the higher harmonics, a weak collision to less amplitude in the higher harmonics. However, a social explanation, that gender-related differences in spectral slope are due to women and men using their vocal organs in a different manner, leads to the same outcome. Nonetheless, the spectral properties of female and male voices can be examined irrespective of whether their genesis is anatomical or social.

Female voices have been shown to be more breathy than male voices (e.g. Henton & Bladon 1985, Klatt & Klatt 1990, Nittrouer, McGowan, Milenkovic & Beehler 1990, Günzburger 1991, Södersten 1994, Ní Chasaide & Gobl 1995, Trittin & de Santos y Lleó 1995). Breathiness is defined in two basic ways in such studies. It can be defined through spectral slope, especially the dominance of the level of the fundamental frequency over the second harmonic or the lowest formant. It can also be defined through high-frequency noise. The inclusion of such parameters in female speech synthesis has increased the quality of the synthesis (e.g. Klatt & Klatt 1990, Karlsson 1992).

The relative prominence of the voice fundamental has been observed in a number of phonetic studies of breathy voice as compared with e.g. tense or laryngealized voice (see e.g. Huffman 1985, Ladefoged & Antoñanzas-Barroso 1985). In many comparisons of female and male voices, the level of the fundamental frequency in relation to that of the lowest harmonics, especially the second harmonic, appears to be an essential measure (see e.g. Karlsson 1976, Monsen & Engebretson 1977, Bickley 1982, Henton & Bladon 1985, Kasuya & Ando 1991, Södersten et al. 1991, De Krom 1994). Klatt and Klatt (1990:829) sum up their results of differences in breathiness between female and male speakers: "To the extent that the first-harmonic amplitude is an acoustic correlate of breathiness, females are more breathy than males."

Female voices have a higher spectral level at very high frequencies. Byrne et al. (1994) carried out a comparison of long-term average speech spectra from more than ten languages. They showed that female voices on the average have a higher level at and above 6.3 kHz than male voices. Similar differences have been observed in female and male singers. The spectra of vowels sung by male singers have little amplitude above 3.5–4 kHz, whereas female singers have regions of spectral energy up to 6–8 kHz (Sundberg 1984).

Figure 2 has been drawn to represent the data reported by Byrne and his numerous collaborators (1994:2116, Table II, normalized and averaged by them across samples for 11 languages). In their study, based on normalized and averaged dB values for female and male subjects at one third-octave bands, Byrne et al. (1994) conclude that in general the differences between women and men are rather small, within 2 dB. However, they fail to interpret their well-collected and interesting data in a framework that would allow further gender-related differences to be revealed.

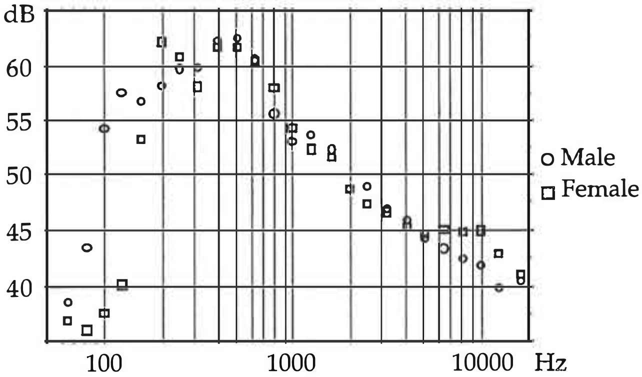


FIGURE 2 Averaged LTA spectrum values from 11 languages separately for female and male speakers. Data from Byrne et al. (1994).

In addition to showing higher amplitudes for female speakers than for male speakers above 5 kHz, discussed above, the figure shows a very clear difference in the levels of the fundamental and the adjacent harmonics, as interpreted on the basis of the obvious fundamental and its multiples (or the bands closest to the multiples). In female speakers, the level of the fundamental (at 200 Hz) is 62.2 dB, the level of the second harmonic (at 400 Hz) is 61.7 dB, and those of the third (at 630 Hz) and fourth (at 800 Hz) harmonics 60.3 dB and 58.0 dB, respectively. In male speakers, the level of the fundamental (at 125 Hz) is 57.7 dB, the level of the second harmonic (at

250 Hz) is 59.7 dB, and those of the third (at 400 Hz) and fourth (at 500 Hz) harmonics 62.4 dB and 62.6 dB, respectively. The trend is evident: female speakers have a spectrum dominated by the fundamental frequency, male speakers a spectrum dominated by the first formant region.

The present study explores spectral differences between female and male speakers and singers. Comparisons are made between dysphonic voices, between soft, normal and loud voices as well as between supported and unsupported and between covered and open singing voices.

4 PROCEDURE

4.1 Research questions

The present study examines voice variation. Voices vary immensely both across and within speakers and singers. The subjects in this study are Finnish, which adds to the knowledge base about voice in different cultures. The basic approach is acoustic, with long-term average spectrum analysis as the data collection method. Therefore, the research questions concern the spectral properties of voice.

Several questions are addressed. The main research question concerns the acoustics of voice.

1 What are the spectral properties of voice?

Four more detailed questions are asked. The questions cover major areas of voice variation: speaking voice and singing voice, normal voice and dysphonic voice, female voice and male voice.

1.1 What are the spectral properties of dysphonic voices? How do groups of dysphonic speakers differ spectrally? How do groups assembled on the basis of diagnosis (laryngeal cancer, vocal nodules, paralysis of the recurrent nerve) and perceptual evaluation (slight, moderate and severe dysphonia) differ spectrally?

1.2 How are the spectral properties of voice related to vocal intensity? What kinds of differences are there in the spectra of soft, normal and loud voices?

1.3 What are the spectral properties that characterize register transition and various voice types in singing? How do supported vs. unsupported and covered vs. open voices differ? How do perceptual ratings of support relate to the spectral properties of support?

1.4 How do the spectra of female and male voices differ?

The answers to these questions supply indirect information about voice production. As much is known about the relation between voice production and voice acoustics, the results can be related to voice production. The adduction–abduction continuum is a central explanatory framework. The second group of research questions deal with the relationship between production and acoustics.

2 What do the spectra of various voices reveal about voice production? How can the spectral observations be related to the adduction–abduction continuum ?

Long-term average spectrum (LTAS) analysis is an established method in the description of voice. However, the interpretation of LTAS spectra has not been standardized. The present study critically evaluates the LTAS method. The third group of research questions concerns the methodology of voice acoustics.

3 What are the advantages and disadvantages of the long-term average spectrum analysis method? What are the essential spectral measures for describing voice variation?

In sum, the research questions deal with the spectral properties of voice, the relationship between voice production and voice acoustics and the long-term spectrum analysis method.

4.2 Material

4.2.1 Dysphonia

The material for the acoustic measurements of dysphonic voices comes from three sources. The first contains speech samples from 128 normal and dysphonic subjects, reported by Sonninen and Hurme (1982). The second source contains speech samples from 40 normal and dysphonic subjects, reported by Sonninen and Hurme (1984) and Hurme & Sonninen (1985a). The third source contains speech samples of 5 individuals with dysphonia, reported by Hurme and Sonninen (1985b). All subjects were native speakers of Finnish. The recordings were made at the Central Hospital of Central Finland by Aatto Sonninen.

By applying stricter criteria than in the three studies mentioned above (Sonninen & Hurme 1982, Hurme & Sonninen 1985a, Hurme & Sonninen 1985b), 87 speech samples were chosen for the present study. Children under 18 were not included and neither were deaf subjects nor rhinolalia cases. Unfortunately, the application of these criteria meant discarding the majority of subjects considered as normal in the previous studies. Consequently, no normal or control group could be formed.

The subjects have all been examined and diagnosed by Aatto Sonninen. They have been divided in four diagnostic groups. In the case of three groups the diagnosis is rather precise (to be described below) whereas the fourth group consists of voice samples from individuals with less severe conditions: functional dysphonia (both habitual and ”ponogenea”), hyperfunctional dysphonia, hypofunctional dysphonia, endocrinological dysphonia, papilloma and sulcus glottidis. The fourth group was subdivided into three groups by Aatto Sonninen, who evaluated the voices as slightly dysphonic, moderately dysphonic or severely dysphonic. The subdivision is based on perceptual evaluation, not on diagnosis. Six groups result (abbreviations are used in presenting the results of measurements):

1 Slight dysphonia (Sl).

2 Moderate dysphonia (M).

3 Severe dysphonia (Sv).

4 The nodule (N) group: individuals with vocal nodules and polyps.

5 The paralysis group (P): individuals with paralysis of the recurrent nerve.

6 The carcinoma group (C): individuals with cancer in the laryngeal region (mainly cancer in the vocal folds but also one case of cancer in the pharynx and one in the epiglottis).

Table 3 presents the groups of dysphonic speakers and gives the total number of speakers and the number of female and male speakers in each group. The mean age of speakers in each group is also given. The abbreviations after group description will be used below in chapter 5.1.

TABLE 3 Groups of dysphonic speakers. The total number of speakers (n), the number of female (f) and male (m) speakers and the mean age in years of speakers in each group.

Group	n	f	m	Age	Description
1	32	20	12	40	Slight dysphonia (Sl)
2	13	6	7	42	Moderate dysphonia (M)
3	18	10	8	45	Severe dysphonia (Sv)
4	4	3	1	40	Nodules, polyps (N)
5	13	10	3	54	Paralysis of vocal folds (P)
6	7	1	6	61	Cancer in laryngeal region (C)
Total	87	50	37	45	

The groups do not have equal numbers of subjects; groups 4 and 6 are small. While both female and male speakers are well represented in the material, their number in small groups can be as low as one. The average age of the subjects is 45 years. Groups 5 and 6 differ from the other groups in comprising older subjects. There is much variation in age in all groups: the standard deviation is highest in group 1 (17.3) and lowest in group 6 (7.5). The age of the subjects ranged from 20 to 82 years.

The subjects read a passage of 30–50 seconds duration from a Finnish children's book. The duration varied with their rate of speech. Their performance was registered on a TEAC tape recorder (tape speed 19 cm/sec) via an AKG CE10 microphone.

4.2.2 Vocal intensity

Ten Finnish university students with no known history of voice disorder read a passage lasting about 30–50 seconds (depending on rate of speech) from the same Finnish book for children as in the dysphonia material. The age of the subjects was 20 to 34. Five of them were females, five males. The subjects read the passage three times at three sound pressure levels. With the aid of a decibel meter placed in front of them at a distance of one meter they first read the passage at 55 dB (soft), then at 65 dB (normal) and then at 75 dB (loud). The decibel meter was set for the A curve. The recordings were made in a sound-insulated room at the University of Jyväskylä. An AKG D–202 microphone, placed at a distance of one meter from the speaker, and a REVOX A 77 tape recorder (tape speed 19 cm/sec) were used. (Hurme 1980a, 1980b.)

To examine the relationship between vocal intensity and fundamental frequency, the average fundamental frequency (F0) in each sample was analyzed by Signalyze (version 2.45; Keller 1992); FFT Comb was used as the algorithm. Table 4 shows the mean fundamental frequency for all subjects during the three tasks.

TABLE 4 Mean fundamental frequency in the samples read at 55 dB, 65 dB and 75 dB; sex of subjects is also indicated.

Subject	55	65	75	Sex
1	209	230	257	female
2	197	212	258	female
3	198	203	234	female
4	175	202	231	female
5	180	198	237	female
6	99	102	108	male
7	107	115	121	male
8	86	99	107	male
9	94	101	123	male
10	125	136	167	male

The mean of the average fundamental frequencies computed from the samples produced at 55 dB is 192 Hz for female and 102 Hz for male subjects. At 65 dB the averages are 209 and 111 Hz, respectively. In speech samples produced at 75 dB the female group average is 243 Hz and the male group average 125 Hz. Thus when comparing average F0 in soft, normal and loud voice, female subjects show the series 192 Hz, 209 Hz and 243 Hz and male subjects 102 Hz, 111 Hz and 125 Hz. Such results are in line with other measurements: it is well known that fundamental frequency increases with loudness (e.g. Dieroff & Siegert 1966, Reimers Niels & Yairi 1987, Gramming et al. 1988, Sonninen et al. 1988).

4.2.3 Supported and unsupported singing

Eight Finnish professional or semiprofessional singers trained in classical singing performed several tasks in supported and unsupported voice. The singers were not helped in defining support and lack of support. Thus, each singer chose him/herself how to sing in a supported voice and in an unsupported voice. Clearly, what is meant by support in this study is how the singers who took part in the study understood support and lack of support.

The singers were registered in a quiet room at the Department of Communication of the University of Jyväskylä, using a REVOX A 77 tape recorder and an AKG D–202 microphone at a distance of 30 centimeters. Several physiological measurements were made simultaneously by means of a twin-channel electroglottograph and a pressure transducer (see Sonninen et al. 1994). The following signals were registered: acoustic signal, electroglottographic signal, vertical laryngeal position, and subglottal pressure.

The tasks that the singers performed were the production of sustained vowels, messa di voce, melodic figure and a song. In all, they produced 170 samples. In this study, only a subset of the material consisting of 64 samples was included: the task in which the singers were instructed to produce a long sequence of [pɑ]-syllables first with support and then without support. The singers were instructed to sing with attempted constant intensity and pitch at a comfortable vocal intensity and with a pitch above and below the subjective register transition area (which is usually located between C4–G4 or 262–392 Hz). In this study, the syllable sequences produced above the transition area are called high and those below low.

The voice class of the singers ranged from soprano to baritone. Two of the male subjects were baritones: subject 1 with a voice range of A2–A4 and subject 3 with a range of E2–G4. The third male subject (subject 2) was a tenor, with a voice range of D3–D5. All the female singers can be classified as sopranos. Their voice ranges were as follows: subject 4 D3–D6, subject 5 A#3–A#5, subject 6 E3–D6, subject 7 D3–A#5 and subject 8 G3–C6.

The fundamental frequency in the syllable sequences was analyzed by Signalyze (version 3.12, Keller 1994); FFT Comb was used as the algorithm. A narrow-band spectrum was first computed to help narrow down the possible range of F0. Table 5 shows the mean fundamental frequency for all subjects during the four tasks with sustained phonation (low supported, high supported, low unsupported, and high unsupported). The abbreviations Sup and Un are used for supported and unsupported voice. The musical note closest to the frequency is also shown. The Hz values for F0 obtained by computing average F0 with FFT Comb are only approximations of the real F0, but they are precise enough for the present purpose: to locate the notes produced on the musical scale.

On the whole, the sequences were produced below and above the register transition area (about C4–G4 or 262–392 Hz). However, subject 6 deviates from the others by having a rather high pitch in the low notes, above the usual register transition area.

TABLE 5 Mean fundamental frequency (in Hz) and the closest musical note in sustained phonations of low and high notes in supported (Sup) and unsupported (Un) voice.

Subject	Voice class	Low, Sup	High, Sup	Low, Un	High, Un
1	baritone	191 (G#)	341 (F4)	195 (G3)	339 (F4)
2	tenor	227 (A#3)	429 (A4)	228 (A#3)	431 (A4)
3	baritone	176 (F3)	371 (F#4)	171 (F3)	364 (F#4)
4	soprano	216 (A3)	647 (E5)	215 (A3)	647 (E5)
5	soprano	238 (A#3)	475 (A#4)	235 (A#3)	472 (A#4)
6	soprano	421 (G#4)	677 (F5)	414 (G#4)	693 (F5)
7	soprano	174 (F3)	601 (D5)	173 (F3)	605 (D5)
8	soprano	185 (F#3)	769 (G5)	184 (F#3)	753 (F#5)

The sound pressure level was not registered during the sessions. As a result, absolute dB values are not available for the description of the overall amplitude of the syllable sequences.

4.2.4 Covered and open singing

A sound recording of a female singer singing in various voice types was made available by Aatto Sonninen. This recording was made in 1953 during a radiographic investigation, the biomechanical results of which have been reported in detail in several publications by Sonninen (1956, 1968), Sonninen et al. (1992) and Sonninen and Hurme (1996). Sonninen (1968) reproduced selected spectrograms and narrow-band spectra of the vowels produced by the singer, but did not present the results of his measurements.

The subject sang a chromatic scale as wide as possible using the vowel [ɑ] both in covered and in open voice. In covered voice she was able to sing from 146 Hz to 1661 Hz (D3–G#6). In open voice the pitch range was 185 Hz to 466 Hz (F#3–A#4). In addition to forte covered and forte open voice, she produced some notes in piano covered and piano open voice. Being a skilful imitator, she also sang some notes in a special loud and open voice type, described by herself as "a young boy's shout". Based on introspection during singing, the subject reported register transition to occur typically between E4–F4 (330–349 Hz).

During the recording session, the sound pressure level for each vowel was measured with a decibel meter at a distance of 75 cm. Thus, absolute dB values for each vowel were available. The tape recorder was of good quality (even though sound registration technology in the early 50s was inferior to present standards). The tape provided by Aatto Sonninen consisted of extracts of 0.5 to 2 seconds duration.

4.3 Methods

4.3.1 Speaking voice

The speaking voice samples have been analyzed by means of long-term average spectrum (LTAS) analysis. There are some differences in the analysis procedure for dysphonic voices and those produced at the three levels of vocal intensity. The differences will be described below.

The acoustic analysis of the speech samples was carried out by means of a 400-channel 2031 Brüel & Kjær FFT analyzer. The speech samples were fed into the analyzer through the built-in 12-bit analogue to a digital converter. The sampling frequency was set to 51.2 kHz in order to obtain a DC to 20 kHz frequency span. With this setting, the spectra presented a 50 Hz resolution over the whole frequency range under investigation. For a subset of the samples the sampling frequency was set to 5.12 kHz to obtain a DC to 2 kHz span and a 5 Hz resolution The Brüel & Kjær 2031 built-in linear averaging process was used in order to compute the long-term spectra. The speech signal was fed into the analyzer through an electronic gate which allowed only voiced segments to pass through (described in detail by Hurme 1980b). In this way unvoiced fricative sounds (especially [s]) were eliminated from the long-term spectra.

In the original analyses of dysphonia (Sonninen & Hurme 1982, Hurme & Sonninen 1985a, Hurme & Sonninen 1985b), the following measurements were made from the LTA spectra: 1, 2, 3, 5, 10 and 14 kHz. In the reanalysis reported here, the measurements cover these points, except the 14 kHz measurement point, which is very unlikely to represent the properties of the signal for reasons that have to do with the recording equipment and environment (Stevens 1985). In the present study, measurement points between 5–10 kHz were included to see if they reveal any systematic differences in the measured phenomena, and also because they have been included in many previous studies (at least up to 8 kHz, e.g. Hammarberg et al. 1980, Hammarberg et al. 1984).

The measurements made from the DC to 20 kHz long-term spectra are described in Figures 3 and 4 (both overleaf). Figure 3 displays actual long-term average spectra up to 10 kHz and the two sets of measurements taken.

The left panel of Figure 3 describes the measurement of dB values at 1 to 10 kHz. The highest peak (invariably located between 0 and 1 kHz) was set to 0 dB. The decibel values of the spectra were then measured in decibels relative to the 0 dB level in 10 steps of 1 kHz; all the values are negative as they were always lower than the reference value. The values above 10 kHz were not measured for the reasons given above. The measurement points are located in the intersections of vertical and horizontal lines.

The right panel of Figure 3 describes the measurement of dB values at two wide-frequency bands. The maximum amplitude level (located between 0 and 1 kHz) was set as the reference point (= 0 dB). In these measurements, the reference is called LL for the dB level at the low frequencies. Maximum levels in the 2–5 kHz area (= LM for level at the mid frequencies) and in the

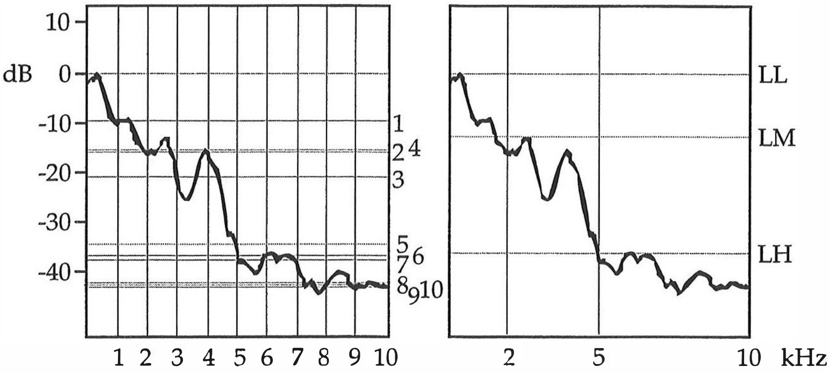


FIGURE 3 A sample LTA spectrum with two sets of measurements: on the left, dB levels at 1 to 10 kHz and, on the right, maximum levels on the 2–5 kHz and 5–8 kHz bands, in both cases in relation to the maximum level.

5–8 kHz area (= LH for level at the high frequencies) were compared to the level at the low frequencies, i.e. LL. In addition, LM and LH were compared. The highest levels within each band are the measurement points.

Further measurements were made from the DC to 20 kHz long-term average spectra to investigate closer the area of maximum amplitude. In Figure 4, only the 0 to 1 Hz area of the LTA spectra is shown, greatly stylized for clarity. The measurements were taken from the screen of the spectrum analyzer, using the built-in measuring function which shows the dB values at steps of 50 Hz. (The magnitude of the step is dependent on the resolution of the analyzer.)

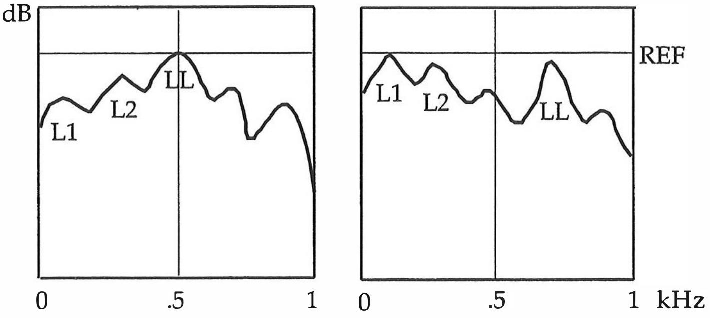


FIGURE 4 Two stylized LTA spectra (only the 0–1 kHz area shown) describing the measurement of the levels of the first harmonic (L1), the second harmonic (L2) and the first formant region (LL).

Figure 4 illustrates two sets of measurements. The levels of the first harmonic (L1) and the second harmonic (L2) were compared with L1 as the reference point. In relation to L1, L2 could be higher or lower; i.e. the slope could be positive or negative. The levels of the first harmonic (L1) and highest peak in the spectrum (LL, typically in the region of the first formant) were compared with LL as the reference point. In relation to LL, L1 could be lower or higher; i.e. the slope could be positive or negative. The left panel in Figure 4 shows an example of a positive slope both when comparing L1 and L2 and when comparing L1 and LL. However, the highest peak in the spectrum was not always located in the first formant area. The right panel in Figure 4 shows an example of a negative slope both when comparing L1 and L2 and when comparing L1 and LL.

When analyzing the samples produced at the three levels of vocal intensity, further measurements were made from the 0 to 2 kHz LTA spectra. Figure 5 shows samples of actual long-term spectra.

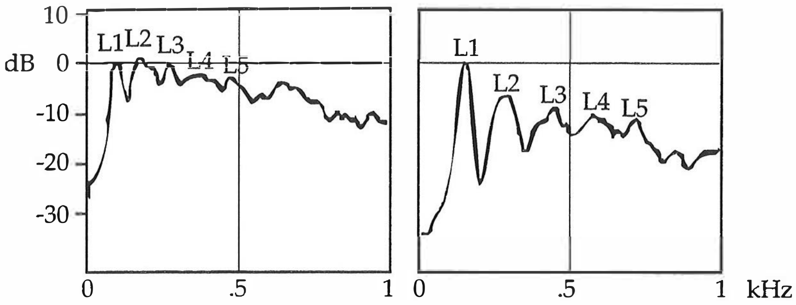


FIGURE 5 Two sample LTA spectra: the levels of the lowest harmonics (L2–L5) measured in relation to the level of the first harmonic (L1). The left panel is an example of a positive slope and the right panel of a negative slope in a comparison of L2 to L1.

The level of the first harmonic (L1) was set to zero dB. The levels of the four harmonics above F0 (L2–L5) were measured in dB in relation to the L1 level. A negative value indicates that the level of the fundamental frequency is higher than that of the higher harmonic; a positive value indicates that the level of the higher harmonic is higher than that of the fundamental frequency. The left panel shows an example of a positive slope when comparing L2 to L1. The right panel shows an example of a negative slope in a comparison of L2 to L1.

The averaging procedure levels the peaks of the harmonics to some extent. If a harmonic could not be identified from the long-term spectrum, the measurement was made at a multiple frequency of the fundamental frequency. However, as the samples in the studies on the singing voice in the present study consisted of sustained vowels only, the peaks were usually prominent.

The measurements were taken directly from the display screen, using the built-in function which shows the dB value at steps of 5 Hz. (The resolution is small as the analyzed frequency range was only 2 kHz.) A permanent record of the LTA spectra was achieved either by taking photographs of the display screen of the analyzer or by drawing the spectra on an x–y plotter. Samples of the long-term average spectra have been published in several articles (Hurme 1980a, 1980b, Sonninen & Hurme 1982, 1984, Hurme & Sonninen 1985a).

4.3.2 Singing voice

Supported vs. unsupported voice

Sequences of the syllable [pɑ] were digitized in a Macintosh computer with an anti-aliasing 8-bit MacRecorder (1991) digitizer at a sampling rate of 22 kHz and using the SoundEdit program. The response of the digitizer is flat (±1 dB) up to 6 kHz. Acoustic measurements were made with the Signalyze program (program level 3.12; Keller 1994). Narrow-band averaged spectra with the bandwidth 30 Hz were computed for two sets of five samples of the vowel [ɑ], chosen at about equal intervals from the beginning and end of the syllable sequences, respectively. Pre-emphasis, which compensates for normal spectral drop-off from low to high frequencies by "boosting" higher frequencies, was off. Log-scale was on, displaying spectral amplitude according to the decibel scale. Only relative measures of the levels in the spectra could be obtained, since the absolute sound pressure level in the recording sessions was not known.

The levels of the spectra were measured in dB at four points: the first harmonic or fundamental frequency (L1), the second harmonic (L2), the highest point in the F1–F2 area (0.5 to 2 kHz, abbreviated LowR) and the highest point in the upper formant area (2 to 4 kHz, abbreviated HighR). The measurements were converted to dB values relative to the highest peak in the spectrum. If the peak was located in the F1–F2 area, this level was assigned zero dB and the other dB values were negative. Figure 6 illustrates the measurements; the reference point is the level at LowR, i.e. the upper reference line.

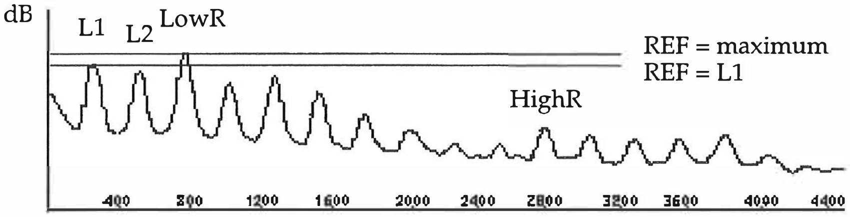


FIGURE 6 A sample averaged spectrum with the four measurement points L1, L2, LowR and HighR (described in text) and two reference lines: the upper for the maximum level to which other levels are compared, the lower for the level of L1 to which the level of L2 is compared.

If the first harmonic was the highest peak, the level of L1 was assigned the value zero dB and the other dB values were negative. The same applies to L2. Thus, the averages calculated for the various groups reveal potential differences in the levels of L1, L2 and LowR. The procedure was adopted to reduce the variation resulting from widely varying pitches in the syllables produced (from about 100 to almost 800 Hz). At high pitches the harmonics are situated, for instance, at 800, 1600, 2400 and 3200 Hz. As there are no other peaks to measure, these peaks represent L1, L2, LowR and HighR. On the other hand, there are many more harmonics at low pitches. Of these, the

the level of the first two harmonics and the highest level in the 0.5–2 kHz area and in the 2–4 kHz area were measured.

In addition to the levels measured relative to the highest peak in the averaged spectrum (usually LowR or L1), a measure with L1 always as the reference point was computed. If L1 is the highest peak, the two measures are the same. However, if L1 is not the highest peak, the L1-L2 measure differs from the measure with LowR or L2 as the point of reference. The L1-L2 difference compares the levels of L2 and L1: if it is positive, the level of the second harmonic is higher than that of the first harmonic. If it is negative, the level of the second harmonic is lower than that of the first harmonic. Figure 6 illustrates the L1-L2 measurements, which were taken in relation to the lower reference line.

The acoustic measurements were complemented with perceptual evaluations of supported and unsupported voice when singing low and high notes. Listening tests where panels consisting of singers, students of singing and voice professionals rated the 64 samples of supported and unsupported voice were administered by Aatto Sonninen and the present writer. The samples were edited digitally from the original tapes to comprise about 5 seconds from the beginning or end of the sample. The raters (n = 63) were asked to evaluate each sample on two visual analog scales: supported vs. unsupported singing and good vs. poor quality. For each sample, the scale was drawn on a response sheet as two segments with supported and unsupported singing and good and poor quality as end points. The rater was asked to draw a mark across the segment, according to his or her evaluation. The segments did not contain any subdivisions. The marks were converted to values from 0 to 100 by measuring them with a ruler. Thus, 100 indicates maximum support and quality, 0 minimum support and quality, i.e. lack of supported in the voice and poor voice quality.

Covered vs. open voice

The recordings by the female singer singing in forte covered and forte open voice as well as other voice types were digitized by means of an anti-aliasing MacRecorder digitizer and a Macintosh computer. The response of the digitizer is flat (±1 dB) up to 6 kHz. In Signalyze (program level 3.12, Keller 1994), each note was extracted from the master file and assigned to its own file, at a sample rate of 22 kHz. In the process, extra noises and clicks (typically from the X-ray apparatus) were eliminated and the occasionally occurring initial palatal glide removed. The result was checked by listening. In Signalyze, an averaged spectrum was calculated for each vowel. First, a spectrogram with a narrow 30 Hz bandwidth was drawn. Then an averaged spectrum was calculated for the vowel: pre-emphasis was off (not ″boosting″ higher frequencies), but log-scale was on (displaying spectral amplitude according to the decibel scale).

From the average spectra four measurements were made in an attempt to capture the essential peaks in the spectrum: decibel values at the fundamental frequency (F0), the second harmonic, the highest peak between 0.5 and 2 kHz, consisting of the lowest two formants F1 and F2, and the

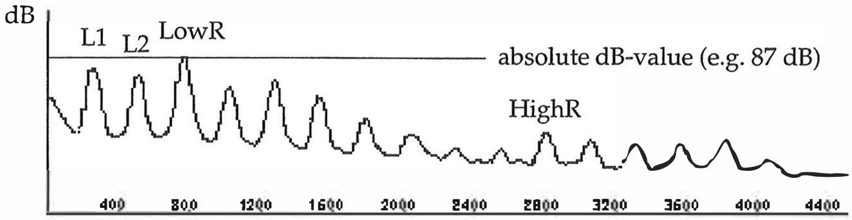


FIGURE 7 A sample averaged spectrum with the four measurement points L1, L2, LowR and HighR (described in text) and an absolute reference line to which the level measurements were compared.

highest peak above 2 kHz, often consisting of formants F3, F4 and F5. Figure 7 illustrates the measurement points. The abbreviations L1, L2, LowR and HighR are used, as in the measurements of supported and unsupported voices.

There are many peaks in the spectra of low notes: of these the two lowest harmonics (L1 and L2) and the LowR and HighR resonances were chosen. In the spectra of high notes (C5 or 523 Hz and above) the first harmonic is in the LowR resonance area. The second harmonic is in the LowR resonance area in notes at C4 (262 Hz) and above. The spectra of very high notes are more straightforward: only the four lowest harmonics can be measured. For instance at D#6 (1245 Hz) the fundamental frequency is at 1245 Hz, the second harmonic at 2490 Hz, the third harmonic at 3735 Hz and the fourth at 4980 Hz.

To take a closer look at register transition in covered and open voice, a more detailed analysis was carried out in a subset of the samples. In the pitch range F#3–A#4 (185–466 Hz), the amplitude levels of harmonics L1, L2, L3, L4, L5 and L6 were measured both in covered and open voice.

The fundamental frequency of each note was measured from the harmonics (by measuring e.g. the 8th harmonic and dividing the frequency by 8). This is not a very precise method, but the purpose was simply to check that the correct note had been measured. (A more detailed analysis of F0 in these samples showed minor and non-systematic differences between the intended and the actual fundamental frequency. This analysis is not reported here.)

The measurements of singing voices are reported both in relative and absolute dB values. Figures 6 and 7 illustrate both sets of measurements. Figure 7 describes the case where the measured dB values have been converted to actual dB values (as the dB values were registered for each note during the original investigation). Figure 6 gives dB values relative to the maximum dB value in the spectrum. In this analysis, the other measurement points receive values that are negative in relation to the maximum level. The levels of the two lowest harmonics were also compared by assigning L1 as the reference and comparing L2 to the reference. A positive dB value indicated that L2 is higher than L1, a negative dB value that L2 is lower than L1.

4.4 Statistical treatment and summary of material and methods

The spectral measurements were described by conventional descriptive statistics: the arithmetic mean and standard deviation. The statistical significance of observed differences between groups has been tested by analysis of variance (oneway ANOVA) by comparing the variation between groups to that within groups. The statistical significance of the observed differences is indicated by asterisks (*** indicates $p \leq .001$, ** indicates $p \leq .01$ and * indicates $p \leq .05$). Lack of statistical significance is indicated by ns (= not significant).

When investigating support in singing, voices were both measured acoustically and evaluated perceptually. The acoustic measures L1, L2, LowR and HighR were related to the perceptual ratings by means of Spearman's rank correlation. The correlation coefficient (r) and the statistical significance of the correlation were calculated. Statistical significance is indicated by the same symbols as above.

The measurement data are often presented in scattergrams of group means and standard deviations. When reporting data from covered and open singing, second-order regression curves have been computed to reduce variation and describe the patterning of the data points.

Table 6 gives a concise summary of the subjects, their tasks and the measurements taken in the four sections of this study. In addition to the acoustic studies, supported and unsupported voices were also evaluated in a listening test.

TABLE 6 Subjects, tasks and measurements in the present study.

Topic	Subjects	Task & environment	Spectral measurements
Dysphonia	87 dysphonic speakers, 6 groups, female and male Ss	Read a text passage at a comfortable level during clinical examination	1–10 kHz, LL, LM, LH, L1-L2, L1-LL, L2–L5
Vocal intensity	10 normal speakers, female and male Ss	Read a text passage at 55, 65 and 75 dB in an anechoic room, self-monitoring vocal intensity	1–10 kHz, LL, LM, LH, L1-L2, L1-LL, L2–L5
Supported singing	8 professional singers, female and male Ss	Sang [pɑ]-syllables in supported and unsupported voice in low and high notes; physiological variables registered as well	L1, L2, LowR, HighR L1-L2
Covered singing	1 professional singer, female	Sang [ɑ]-syllables in covered and open voice in a rising pitch series; lateral spot radiograms taken as well	L1, L2, LowR, HighR L1–L6, L1-L2

5 RESULTS

5.1 Dysphonia

The results of the long-term average spectrum measurements for the dysphonic voices are presented separately for the three methods of measurement (dB levels at 1–10 kHz in 1 kHz steps, dB levels at low, middle and high frequency bands, and dB levels in the two lowest harmonics). Further comparisons will then be presented.

5.1.1 Spectral levels at nine frequencies

The overall results of the measurements, with all subjects pooled, are shown in Figure 8.

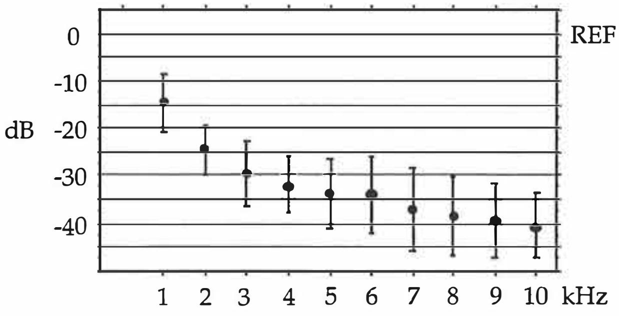


FIGURE 8 Mean dB values at 1 to 10 kHz in relation to the maximum level of amplitude (= REF) in all subjects (n = 87). Means and standard deviation are indicated by dots and whiskers.

The figure retains the co-ordinates of the Brüel & Kjær Narrow-Band Spectrum Analyzer, representing amplitude vertically and frequency horizontally. The point of reference, the maximum level of amplitude, which is located between 0 and 1 kHz, is indicated by REF.

The decrease in dB values is rather steep: about -15 dB from the reference level to the level at 1 kHz, -10 dB from 1 to 2 kHz and -5 dB from 2 to 3 kHz. On the whole, the dB values are rather low (about -35 to -40 dB) at the upper frequencies.

The data points show much variation. In general, there seems to be more variation in the upper than lower frequencies. Analysis of variance (ANOVA) frequently shows greater variance between groups than within groups. At 1 and 6 kHz between-group variance is significantly greater than within-group variance (F=2.9* and F=3.2*). However, at 3 and 4 kHz within-group variance is greater than between-group variance.

Dysphonic voices have been assigned to groups on the basis of diagnosis (nodule, paralysis and cancer groups) and on the basis of perceptual evaluation (slight, moderate and severe dysphonia). Figure 9 shows the mean dB values in each group.

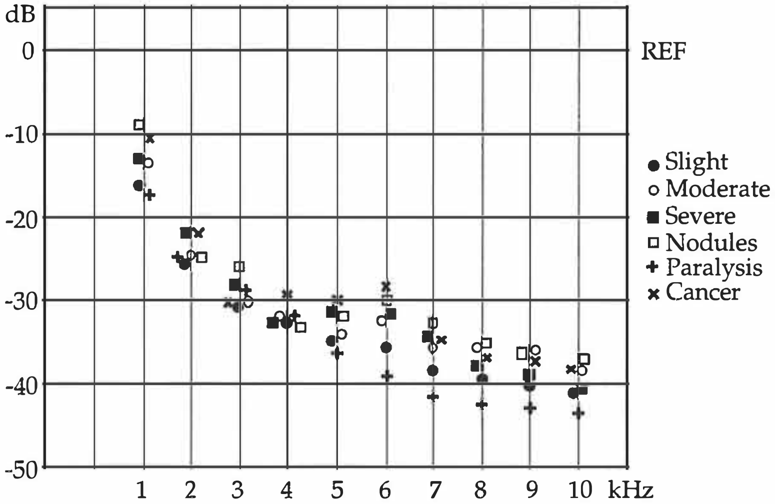


FIGURE 9 Group means of the slight, moderate and severe dysphonia groups and the nodule, paralysis and cancer groups at 1 to 10 kHz with the maximum dB level as the point of reference (REF).

Systematic patterns cannot easily be seen in the data. However, the cancer group and the nodule group often show high dB values, whereas the paralysis group shows rather low dB values.

The data points have been reorganized in Figure 10 so as to highlight the differences between the groups. This figure indicates the positive or negative difference in dB of each group mean at each measurement frequency in relation to the mean for all subjects (always set to zero at all measurement points). In Figure 10 a positive value indicates that the mean for that group is higher than the mean for all subjects, a negative value the opposite.

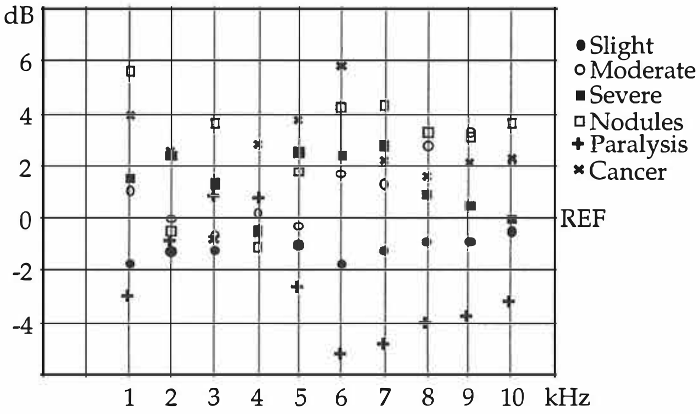


FIGURE 10 The difference between the mean dB level for each voice group and the mean for all subjects (set to zero) at 1 to 10 kHz.

Figure 10 is easier to interpret than Figure 9. For instance, the dB values of the paralysis group are typically lower than the mean for all subjects and those of the cancer group higher. The differences between the group means are significant in a number of cases. The statistically significant differences between the group means shown in Figure 9, and in an altered form in Figure 10, are as follows (calculated by ANOVA):

At 1 kHz: slight dysphonia vs. nodule (F = 1.1*), slight dysphonia vs. cancer (F = 1.1*), severe dysphonia vs. paralysis (F = .9*), nodule vs. paralysis (F = 1.3*), paralysis vs. cancer (F = 1.2*).
At 2 kHz: slight vs. severe dysphonia (F = 1.2*).
At 3 and 4 kHz: none.
At 5 kHz: severe dysphonia vs. paralysis (F = .8*).
At 6 kHz: slight dysphonia vs. cancer (F = 1.1*), moderate dysphonia vs. paralysis (F = 1.2*), severe dysphonia vs. paralysis (F = 1.6**), nodule vs. paralysis (F = 1.0*), paralysis vs. cancer (F =2.0**).
At 7 kHz: severe dysphonia vs. paralysis (F = 1.2*).
At 8–10 kHz: moderate dysphonia vs. paralysis (F = 1.0*, F = .8*, F = 1.2*).

On the frequency scale, the significant differences concentrate at 1 and 6 kHz. A pattern starts to emerge, where the statistically significant differences between group means often involve the paralysis and the cancer groups. There are significant differences between several other groups as well.

By computing a mean for each diagnostic group across all frequencies, an even simpler pattern is obtained. Figure 11 (overleaf) shows the mean difference of each group from the mean for all subjects, with the data from 1 to 10 kHz pooled.

Analysis of variance shows that the variance is much greater between groups than within groups (F = 18.0***). The differences between group means are generally statistically significant, except between the moderate and severe dysphonia groups and between the nodule and cancer groups.

Therefore, the six groups can be reduced to four: the paralysis group, the slight dysphonia group, the moderate/severe dysphonia group and the cancer/nodule group. The first two groups show lower mean dB levels and last two groups higher mean dB levels than the mean for all subjects. The

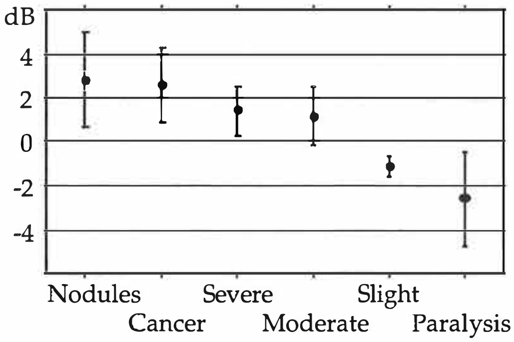


FIGURE 11 Mean difference in dB values in six voice groups at 1–10 kHz compared to the mean of all subjects (which was set as the point of reference at each measurement frequency). Means and standard deviation are indicated by dots and whiskers.

groups that differ most from each other in this analysis are the paralysis group and the cancer/nodule group.

5.1.2 Spectral levels in three frequency areas

To complement the measurements described above, the LTA spectra of the subjects' voices were also measured using a procedure with fewer measurement points. The maximum dB level of the spectrum was set as the point of reference; the reference is invariably located at the low frequencies (0–2 kHz, LL). The maximum dB levels at the middle frequencies (2–5 kHz, LM) and at the high frequencies (5–8 kHz, LH) were then compared to the reference level (= LL). In addition, LM and LH were compared with each other. Figure 12 shows the means and standard deviation for each group.

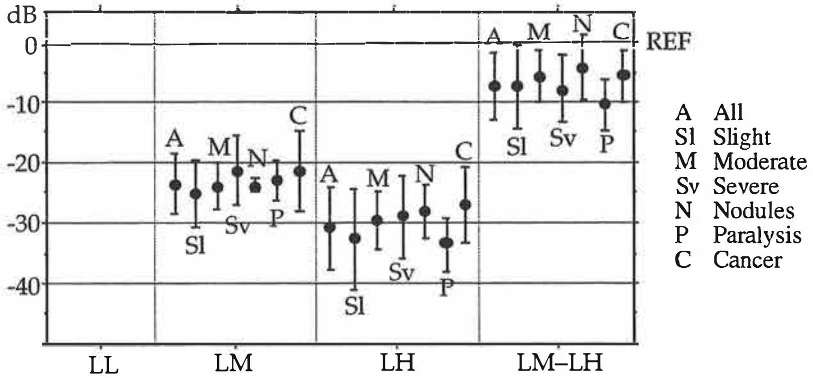


FIGURE 12 Results of LM and LH measurements in relation to LL (= REF) for all subjects and for the voice groups. Means and standard deviation are indicated by dots and whiskers.

In all three measures, ANOVA shows that the variance between groups is greater than within groups. However, statistical significance is not reached at this level of analysis.

A comparison of the groups with each other shows some significant differences. In severely disordered voices and in the cancer group the middle frequency level (LM) is higher than the mean for all groups. In the slightly dysphonic group LM is lower than the mean for all groups. However, only the difference between the slight and severe dysphonia groups is significant (F = 3.0*). At the high frequencies (LH) the nodule and cancer groups have a mean dB level higher than that for all subjects and the slight dysphonia and the paralysis group a mean lower than that for all subjects. The differences between the cancer group and the paralysis group and the cancer group and the slight dysphonia group are significant (F = 6.3* and F = 5.6*, respectively). In the LM–LH comparison, the nodules group and the cancer group deviate from the mean for all subjects in a positive direction and the paralysis group in a negative direction. The paralysis group and the moderate group differ significantly (F = 4.4*).

On the whole, these results point in the same direction as the measurements from 1 to 10 kHz reported above: the cancer and the nodule or severe dysphonia group are above the mean and the paralysis and slight dysphonia group below the mean. The similarity in the results is hardly surprising as the two sets of measurements are related: the latter is a simplified set of the former.

5.1.3 Spectral levels of the lowest two harmonics

The level of the fundamental frequency (L1) is often compared with the level at the low formant region (consisting of F1 and in some cases F2). In the present material this measure does not work well, as L1 can be the maximum level of the spectrum with the result that no clear spectral peak can be seen in the F1–F2 area.

Here the low end of the spectrum was investigated by setting the level of the fundamental frequency (L1) as the point of reference and by comparing the level of the second harmonic (L2) to it. L2 can be either higher than L1, resulting in a positive dB value, or lower, resulting in a negative value. The results of the comparison are shown in Figure 13 (overleaf) with means and standard deviation pooled for each voice group and also for all subjects. An analysis of variance shows that the variance between groups is greater than within groups. However, statistical significance is not reached at this level of analysis.

The level of the second harmonic is lower than the mean for all subjects in the paralysis and slight dysphonia groups. It is close to the mean in the nodule and severe dysphonia groups. In the moderate dysphonia group the level of the second harmonic is slightly higher than the mean for all subjects and in the cancer group clearly higher. There is less variation in the cancer group than in the other groups. The only significant differences are between the cancer group and the paralysis group (F = 3.5*) and between the cancer group and the slight dysphonia group (F = 3.1*).

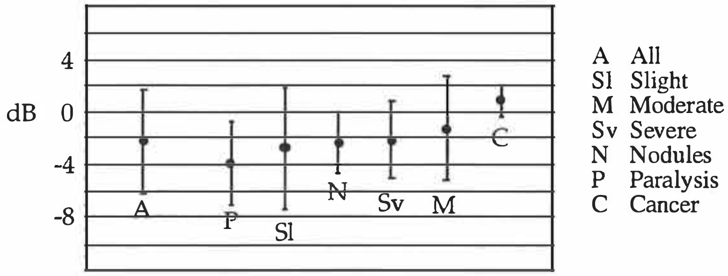


FIGURE 13 Mean dB level of the second harmonic (L2) related to that of the fundamental frequency (L1, set to zero dB) for all subjects and for the voice groups. Means and standard deviation are indicated by dots and whiskers.

The three sets of measurements reported above agree in singling out the paralysis group and the cancer group as diverging most from the mean for all subjects. Levels higher than the mean characterize the cancer group, and levels lower than the mean the paralysis group. Of the other groups, the slight dysphonia group often accompanies the paralysis group, and the nodule and the severe dysphonia groups accompany the cancer group. The position of the moderate dysphonia group is intermediate.

5.1.4 Further comparisons

To compare the long-term average spectra of female and male subjects, the spectra were pooled according to the sex of the speaker. Figure 14 shows the results by means of means and standard deviation, measured at nine frequencies from 1 to 10 kHz.

In general, analysis of variance shows that variance between groups is greater than within groups. Figure 14 shows that the female subjects tend to have higher dB values than the male subjects. The difference is statistically significant at 1 kHz (F = 5.1*) and at 3 kHz (F = 33.7***). The spectral slope decreases faster in the male than female subjects and stays at a lower level at the higher frequencies. However, the differences in the means are not large and there is much variation between individuals, as shown by the standard deviations.

An analysis of the differences of the mean dB values of the female and male subjects has also been carried out in each diagnostic group. However, such comparisons are not very reliable, as there are very few subjects in some groups: only one female subject in the cancer group and one male subject in the nodule group. Nevertheless, the comparison (which is not reported here in detail) shows that the female–male differences within the voice groups are not consistent except at the middle frequencies (especially at 3 kHz in the 1–10 kHz measurements and at LM, i.e. the highest value in the 2–5 kHz region), where the female subjects have higher values.

The level of the second harmonic (L2) is on the average 2.7 dB lower than that of the first harmonic (L1) in the female subjects and 1.5 lower

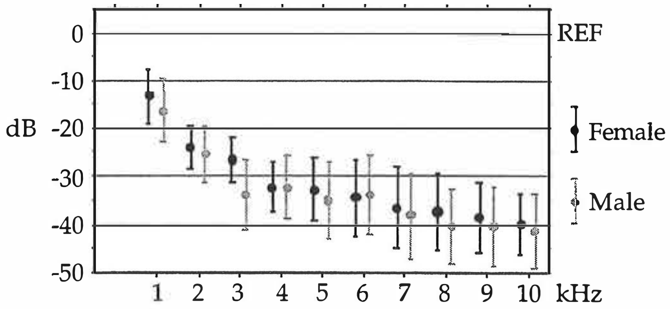


FIGURE 14 Mean dB values and standard deviation, indicated by dots and whiskers, in the long-term spectra of female (n = 50) and male (n = 37) subjects.

in the male subjects; the difference is not statistically significant. However, this difference runs counter to the differences between the female and male subjects described above in both the 1–10 kHz measurements and in the LM and LH measurements. The L1-L2 differences show that the spectral slope is steeper in the female than male subjects, while the 1–10 kHz as well as the LM and LH measurements show that the spectral slope is steeper in the male subjects.

A further comparison of the dysphonic voices was made by forming two new groups: the hypofunctional and the hyperfunctional. These groups consist of individuals in the slight, moderate and severe dysphonia groups. Their diagnosis contained the term hypofunctional (n = 27) or hyperfunctional (n = 15), occasionally accompanied by other characterizations. The dB levels at 1 to 10 kHz for the two groups are shown in Figure 15.

Analysis of variance shows in general greater variance between groups than within groups. None of the differences between group means for the hypofunctional and hyperfunctional voices are statistically significant. The trend, however, is for the hypofunctional voices to show higher mean values at and above 3 kHz and the hyperfunctional voices at 1 and 2 kHz.

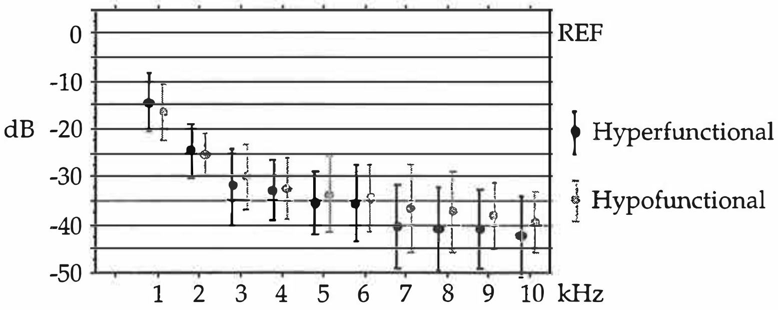


FIGURE 15 Mean dB values and standard deviation, indicated by dots and whiskers, in the long-term spectra of subjects diagnosed as hypofunctional (n = 27) and hyperfunctional (n = 15).

The level of the second harmonic (L2) is on the average 3.4 dB lower than that of the first harmonic (L1) in the hypofunctional subjects and 1.2 dB lower in the hyperfunctional subjects. However, the difference is not statistically significant.

The subjects vary considerably in age (see Table 3 on page 54). Analysis of variance shows no significant differences in any of the acoustic measures between the younger and older subjects. Differences in group means are small – generally smaller than between the female and male subjects and between the hypofunctional and hyperfunctional voices. Thus, age-related acoustic differences do not emerge in the present material.

5.1.5 Synopsis of the results on dysphonia

The results reported above are summarized in Figure 16 and Table 7. Figure 16 describes the main lines of the results for selected groups by six spectral measures. Table 7 summarizes the significant differences between the diagnostic groups at three areas of the spectrum.

Figure 16 shows the direction of the results for the "extreme" diagnostic groups, the paralysis group (P) and the cancer group (C), here comprising the nodule group, hypofunctional (–) and hyperfunctional (+) voices and female (F) and male (M) speakers. The measurements at 6 to 10 kHz have been collapsed, as the dB values at these measurement points almost invariably behave in the same manner. The levels of the second (L2) and first harmonics (L1) are also compared in the figure. The point of reference is L1 in the L1-L2 comparisons (indicated by REF:L1) and the maximum dB level (LL, indicated by REF:LL) in the other measurements. However, the vertical axis is arbitrary, not a dB scale. It should also be noted that Figure 16 includes some statistically nonsignificant observations in addition to significant differences; the intent is to illustrate the trends in the results.

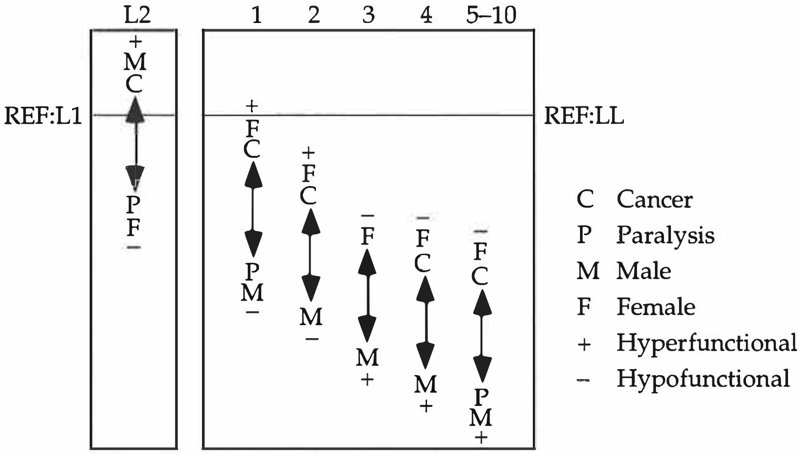


FIGURE 16 Schematic comparisons of groups. The points of reference are L1 (REF:L1) and maximum spectral level (REF:LL).

Before discussing the trends visible in Figure 16, Table 7 is presented. It combines three statistically significant differences in spectral slope: the level of the second harmonic to that of the first harmonic (L1-L2), the level at 1 and 2 kHz (1 & 2 kHz) and the level at the higher frequencies (3+ kHz). The table shows the groups which have higher spectral amplitude (indicated by ▲) and the groups that have lower amplitude (indicated by ▼) in a certain frequency area. Table 7 also shows in parentheses some non-significant but systematic trends in the results.

TABLE 7 Statistically significant (and in parentheses non-significant but systematic) differences in spectral slope in the L1-L2, 1 & 2 kHz and 3+ kHz measures. The groups indicated by ▲ have higher spectral amplitude, those by ▼ lower amplitude.

	L1-L2	1 & 2	3+
▲	cancer (male)	severe, nodule (hyper) female	cancer, severe, nodule (hypo) female
▼	paralysis, slight (female)	paralysis, slight (hypo) male	paralysis, slight (hyper) male

Figure 16 and Table 7 indicate that the level of the second harmonic (L2) compared to the L1 reference point is relatively high in the cancer group, in the male subjects and in the hyperfunctional voices and relatively low in the paralysis group, in the female subjects and in the hypofunctional voices. In the breathy–creaky dimension, the male subjects, the cancer group and the hyperfunctional group appear to lean toward creaky voice, whereas the female subjects, the paralysis group and the hypofunctional group appear to lean toward breathy voice. These groups appear to share some qualities.

In the 1–10 kHz measurements, the mean dB values are relatively high in the cancer group and relatively low in the paralysis group, though at some frequencies the means for these groups do not differ much from the mean for all subjects. In other words, the overall spectral slope is steeper in the paralysis group and shallower in the cancer group. The hyperfunctional voices behave like the cancer group at 1 and 2 kHz, but at the higher measurement points the hypofunctional voices behave like the cancer group. The relatively high dB values in hypofunctional voices at 3, 4 and 5–10 kHz may be an indication of noise or turbulence (at least in some voices). The female and male subjects differ in a consistent way: mean female dB values are relatively high and male values relatively low. In other words, the male subjects show a steeper spectral slope than the female subjects, at least some of whom may have turbulence in their voice. In sum, the spectral slope is steeper in the paralysis group and in the male voices than in the cancer group and the female voices. The hypofunctional and hyperfunctional groups behave in a dual manner, as described above.

5.2 Vocal intensity

The results of the long-term average spectrum measurements of the loud, normal and soft voices are presented in a similar manner as those of the dysphonic voices. The results are presented separately for the three methods of measurement: dB levels at 1–10 kHz in 1 kHz steps, dB levels at the low, middle and high frequencies, and the dB levels of the lowest harmonics.

5.2.1 Spectral level at nine frequencies

The long-term average spectra were measured from 1 to 10 kHz in steps of 1 kHz in relation to the highest dB value (LL) located between 0–1 kHz. Figure 17 gives the means and standard deviation for each of the steps, pooled for all subjects and for the three loudness conditions (loud, normal, soft).

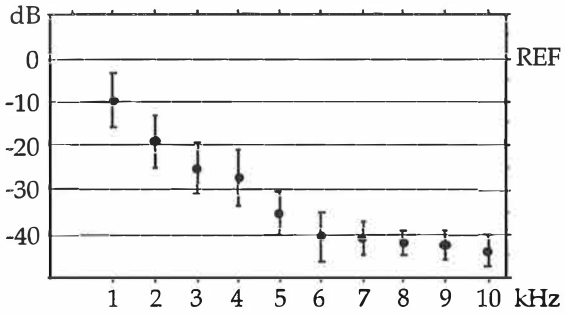


FIGURE 17 Means and standard deviation, indicated by dots and whiskers, of spectral data points in dB for all subjects (n = 10) with the loudness conditions pooled, measured in steps of 1 kHz from 1 kHz to 10 kHz in relation to the highest dB value (REF).

The mean values decrease by about 10 dB/kHz from the maximum value to 2 kHz. From 2 to 6 kHz the decrease is smaller, and the mean dB values from 6 to 10 kHz differ little. The mean values at 6–10 kHz are more than 40 dB lower than the maximum. There is a considerable amount of variation.

Part of the variation may be explained by the loudness conditions. The means of the measurements for all subjects (both female and male) in the three loudness conditions are shown in Figure 18.

The mean dB values clearly differ across the three loudness conditions in the 1–4 kHz area. Analysis of variance shows higher variance between groups than within groups in the 1–3 kHz area (at 1 kHz: F = 19.2***, at 2 kHz: F = 24.5***, at 3 kHz: F = 6.1**). The decrease in the dB values in

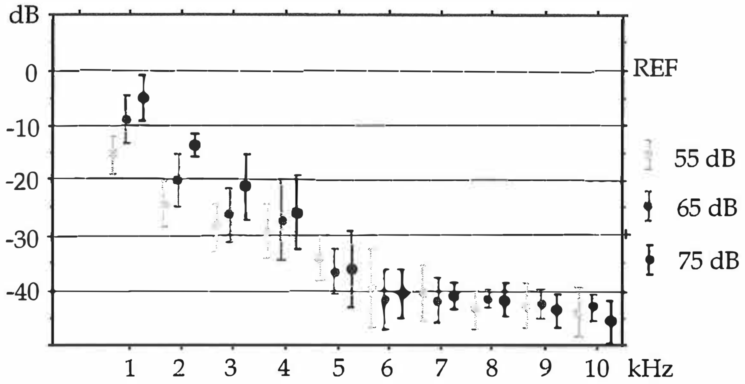


FIGURE 18 Means and standard deviation, indicated by dots and whiskers, of spectral data points in dB for all subjects (n = 10) in the three loudness conditions (55, 65 and 75 dB), measured in steps of 1 kHz from 1 kHz to 10 kHz in relation to the highest dB value (REF).

relation to the point of reference, i.e. spectral slope, is steeper in soft voice than in loud voice, with normal voice in-between. Analysis of variance shows that in the 1–3 kHz area the differences in the means are statistically significant between the 55 dB and 75 dB conditions. Likewise, the differences in the means of the 55 dB and 65 dB conditions as well as in the 65 dB and 75 dB conditions (except at 3 kHz) are significant. At 4 kHz the differences in the means are in the same direction as in the 1–3 kHz area, but they are not significant. In the 5–10 kHz area the differences are smaller and non-significant; no pattern can be seen here in the dB values.

A comparison of the means for female and male subjects in the three loudness conditions shows some significant differences. In soft voice (55 dB), the mean is higher in the male subjects than female subjects at 4 kHz (F = 5.7*). In loud voice (75 dB) female dB values are higher at 1 and 5 kHz (F = 86.9*** and F = 7.0*, respectively). Normal voices (65 dB) show no significant differences between female and male subjects. Looking at the direction of the differences between female and male speakers across the loudness conditions, only one systematic difference can be observed: at 1 kHz the female subjects show higher dB values than the male subjects in all conditions. The difference is significant in loud voice as described above and close to significance in normal voice (F = 5.2, p = .052ns).

5.2.2 Spectral levels in three frequency areas

Comparisons of maximum dB levels in three frequency areas (low, middle and high) are presented in Figure 19 (overleaf). The maximum dB values at 2–5 kHz (LM) and the maximum dB values at 5–8 kHz (LH) were compared to the overall maximum dB level, the LL point of reference. In addition, LM and LH were compared with each other.

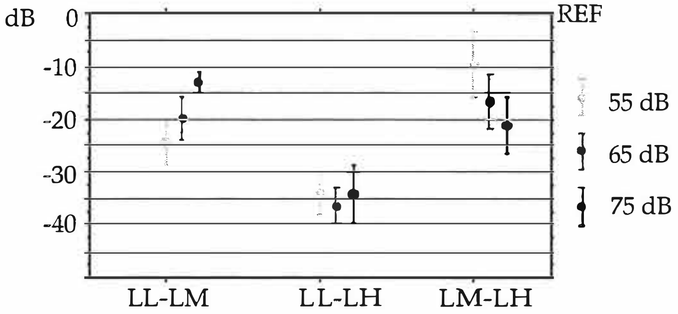


FIGURE 19 Means and standard deviation (dots and whiskers) of dB level comparisons LL–LM, LL–LH, LM–LH for all subjects (n = 10) in the three loudness conditions (55, 65 and 75 dB).

The mean dB values clearly differ across the three loudness conditions in the LL–LM measures. ANOVA yields a significantly higher variance between groups than within groups (F = 26.7***). It can be seen that LM (level at 2–5 kHz) increases in relation to LL (the reference point) with increasing loudness, i.e. the spectral slope becomes less steep. All the differences between the means in the three loudness conditions are significant.

The LL–LH (reference vs. 5–8 kHz) relation shows no systematic trend. On the other hand, the LM–LH (2–5 vs. 5–8 kHz) difference increases with increasing loudness, mirroring the LL–LM behavior. This is clearly a consequence of the increase in LM level with loudness, as the LH level remains about the same in the three loudness conditions.

The LM and LH measurements are not reported here separately for female and male subjects, as none of the differences between the means are statistically significant and no pattern can be discerned in the directions of the differences. The result is expected, as few systematic differences were observed between the female and male subjects in the measurements at the nine frequencies.

5.2.3 Spectral levels of the lowest harmonics

The dB levels of the lowest four harmonics (L2–L5) were measured in relation to the level of the first harmonic (L1). The results are given in Figure 20 on a logarithmic frequency scale with real Hz values, i.e. arranged according to the fundamental frequency and its multiples.

The spectral slope appears steepest in the 55 dB loudness condition and relatively level in the 75 dB condition. However, it is difficult to discern differences in the steepness of the spectral slope, as the fundamental frequency (and hence the location of the data points in the Hz scale) is different in the female compared to male subjects. The data for the male subjects cluster around the low frequencies (mean F0 a little above 100 Hz, see Table 4 on page 55), whereas those for the female subjects extend higher in the frequency scale (mean F0 around 200 Hz).

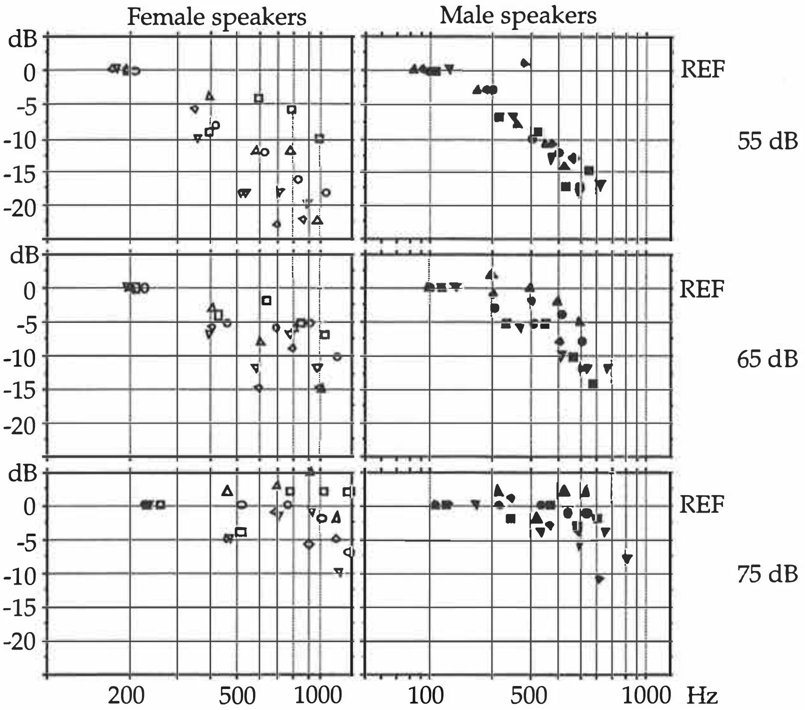


FIGURE 20 Measurements of dB levels of L2 to L5 in relation to L1 (set to zero dB) in the three loudness conditions (55, 65 and 75 dB) for each subject (n = 10), arranged on a logarithmic frequency scale. Female subjects indicated with hollow symbols, male with filled.

To eliminate the effect of the fundamental frequency differences, the measurements of the dB levels of the lowest harmonics were converted to an equidistant scale, in accordance with e.g. Fant (1973). Figure 21 shows the mean dB levels of harmonics 2 to 5 with that of the first harmonic as a point of reference (indicated by REF) and with the subjects and loudness conditions pooled.

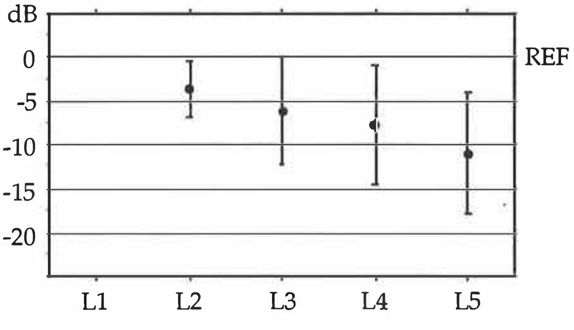


FIGURE 21 Means and standard deviation, indicated by dots and whiskers, of dB values of harmonics 2 to 5 (L2–L5) in relation to that of the first harmonic (L1, indicated by REF), measured from speech samples produced by all subjects (n = 10) in all loudness conditions.

Compared to the level of L1, the slope is downwards; the decrease is of the order of 3 dB per harmonic. However, there is much variation.

Part of the variation is explained by the loudness conditions. Figure 22 shows the mean dB values of harmonics 2 to 5 against that of the first harmonic as a point of reference in soft, normal and loud voice. The scale is equidistant, as in Figure 21.

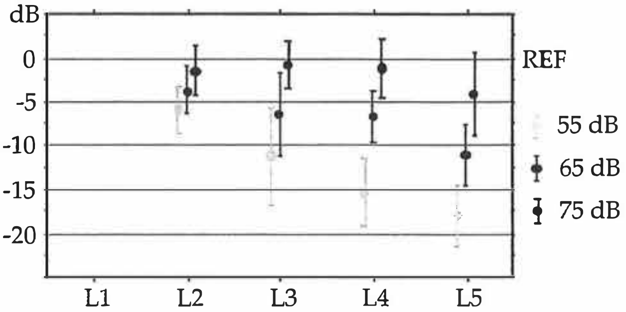


FIGURE 22 Means and standard deviation, indicated by dots and whiskers, of dB values of harmonics 2 to 5 (L2–L5) in relation to that of the first harmonic (L1, indicated by REF) measured from speech samples produced by all subjects (n = 10) in the three loudness conditions (55, 65 and 75 dB).

Even though there is quite a lot of interindividual variation, a clear trend can be seen. Loud voice is associated with a relatively shallow slope; in other words, the level of partials 2 to 5 is almost equal to that of the fundamental frequency. On the other hand, soft voice is associated with a steep spectral slope: the fundamental frequency dominates and the level of partials 2 to 5 is relatively weak. Normal voice lies in between. Compared with the first harmonic, each harmonic is on the average about 1 dB lower in loud voice, 3 dB lower in normal voice and 5–6 dB lower in soft voice.

In all the measures, variance between groups is higher than within groups (L2: F = 6.8**, L3: F = 13.7***, L4: F = 42.6***, L5: F = 31.4***). At L4 and L5 the differences are significant between all conditions, at L3 between the loud and soft conditions as well as between the loud and normal conditions, and at L2 between the loud and soft conditions.

Part of the variation is explained by female–male differences. Figure 23 compares the female and male subjects in the three loudness conditions. The L2–L5 measurements are given on an equidistant scale.

In soft voice, the overall spectral slope is steeper in the female subjects than in the male subjects. The difference is significant at L3 and L5 (L3: F = 6.6*, L5: F = 24.5**). In normal voice, the slope is again steeper in the female than male subjects (except at L4), but none of the differences are significant. In loud voice, the female subjects show higher mean dB values at L3 and L4, and about the same at L5. However, L2 in loud voice is lower in the female than male subjects, but the difference is not statistically significant. Summing up the trends in the results on spectral slope, the female subjects show the lowest means (in the 55 dB condition) and the highest means (in the 75

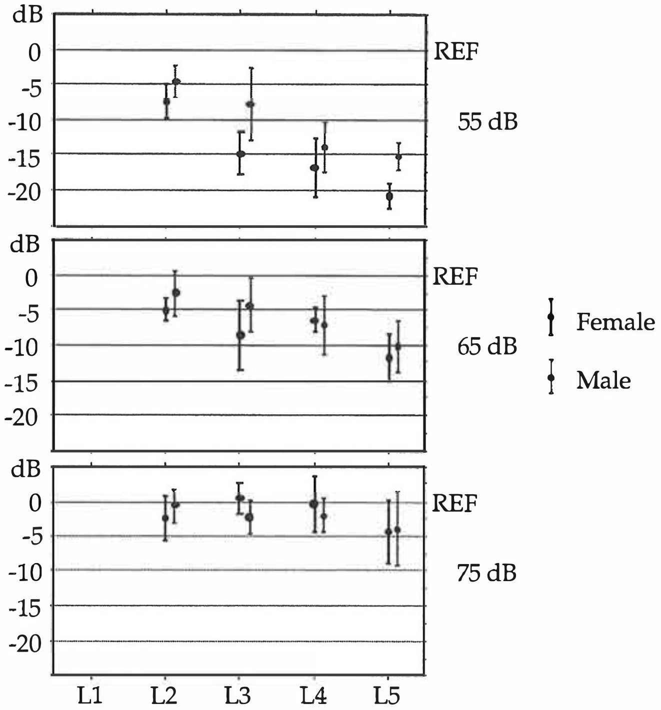


FIGURE 23 Means and standard deviation, indicated by dots and whiskers, of dB levels of harmonics 2 to 5 (L2–L5) in relation to the fundamental frequency (L1, indicated by REF) in samples produced in the three loudness conditions (55, 65 and 75 dB) by female (n = 5) and male subjects (n = 5).

dB condition). The male subjects show less variation in the mean dB values in the three conditions than the female subjects.

A comparison of L2 to L1 gives consistent results: the level of the second harmonic is on the average lower in the female than male subjects. The L1-L2 difference is presented in Table 8 (overleaf). In addition to this measure, the difference between the first harmonic (L1) and the over-all amplitude maximum (LL) was measured, taking L1 as the point of reference. LL was usually in the F1–F2 area, but in a number of cases L1 had the highest amplitude. If L1 is the highest level in the spectrum, the dB value for LL is set to zero. Table 8 shows the mean differences in the amplitude levels of LL and L1 as well as L2 and L1.

A comparison of the levels of the lowest two harmonics with female and male subjects pooled repeats the pattern already seen in Figure 22 (which includes the data presented in Table 8): in relation to L1, L2 is highest in loud voice and lowest in soft voice and in-between in normal voice. As described above, the variance between groups is higher than within groups (L1-L2: F = 6.8**). The difference between the means of loud and soft voice is statistically significant.

In regard to loudness conditions, a comparison of female and male subjects separately yields similar results. Spectral slope as indicated by the

TABLE 8 Mean differences and standard deviation in dB levels of L2 and LL with L1 as the reference point in the three loudness conditions (75, 65 and 55 dB) with female and male subjects pooled as well as separately for female and male subjects.

		L1-L2		L1-LL	
		x	s	x	s
Female & Male	75	–1.5	2.8	1.0	1.6
	65	–3.6	2.7	0.2	0.7
	55	–6.0	2.6	0.1	0.3
Female	75	–2.4	3.2	1.4	2.2
	65	–5.0	1.6	0	0
	55	–7.4	2.4	0	0
Male	75	–0.6	2.4	0.6	0.9
	65	–1.8	3.0	0.4	0.9
	55	–4.6	2.2	0.2	0.4

L1-L2 difference is relatively steep in soft voice and level in loud voice with normal voice intermediate. In the female subjects, ANOVA yields a significant difference between the means of loud and soft voice (L1-L2: F = 5.0*).

A comparison of L1 and L2 across female and male subjects shows consistent differences between the sexes. At high vocal intensity (75 dB), the mean difference in level for all subjects is -1.5 dB; the mean for the male subjects is -.6 dB and for the female subjects -2.4 dB. At normal vocal intensity (65 dB), the mean for all subjects is -3.6 dB, with -2.6 for the male subjects and -5.0 dB for the female subjects. At low vocal intensity (55 dB), the mean for all subjects is -6.0, with -4.6 dB for the male subjects and -7.4 dB for the female subjects. Thus, the female subjects consistently have a steeper L1-L2 slope in all the loudness conditions. However, the differences are not statistically significant.

Comparison of the highest dB level (LL) to the level of the fundamental frequency does not show systematic differences in the loudness conditions. This is due to the fact that the level of the fundamental frequency often is the highest peak in the spectrum, especially in female voices. Therefore, it can be concluded that in this material the L1-LL difference is not a measure that differentiates voices, whereas L1-L2 is.

5.2.4 Synopsis of the results on vocal intensity

Table 9 combines three statistically significant differences in spectral slope between loud and soft voice: the levels of the first and second harmonic (L1-L2), the level at 1 and 2 kHz (1 & 2 kHz) and the level at

TABLE 9 Statistically significant differences in spectral slope in three measures (L1-L2, 1 & 2 kHz and 3+ kHz). The groups indicated by ▲ have higher spectral amplitude, those with ▼ lower amplitude. Statistically non-significant but systematic trends are indicated in parentheses.

	L1-L2	1 & 2	3+
▲	loud (male)	loud female	loud
▼	soft (female)	soft male	soft

higher frequencies (3+ kHz). In addition to statistically significant differences, the systematic trend observed in the L1-L2 difference between female and male subjects (see Table 8) is included in Table 9. The table shows the groups which have higher spectral amplitude (indicated by ▲) and the groups that have lower amplitude (indicated by ▼) in a certain measure.

The loud and soft voices show unambiguous results. The loud voices have a shallower spectral slope and the soft voices a steeper slope. The female subjects have significantly more amplitude at 1 and 2 kHz than the male subjects. On the other hand, in the male subjects the level of the second harmonic (L2) tends to be higher than in the female subjects.

The results of the 1 to 10 kHz and the L2 to L5 measurements presented above can also be shown graphically for the female and male subjects separately. If the data are arranged according to vocal intensity so that the maximum dB level is placed at 75, 65 and 55 dB, the spectral slope can better be appreciated. This has been done in Figure 24 (overleaf), where the point of reference is the maximum amplitude, and in Figure 25 (overleaf), where the reference point is the level of the first harmonic. It is true that there may be sources of error in these procedures. First, the sound pressure level was not measured during the recording sessions and the actual dB levels are unknown; however, the subjects were instructed to try to speak at the prescribed levels by monitoring themselves using the decibel meter. The second possible source of error applies only to Figure 25: the choice of the first harmonic (L1) as the reference point in the descriptions of the low harmonics (L2 to L5), is not entirely adequate. Figure 20 (on page 77) shows that the maximum dB level is not invariably located on the first harmonic. Nevertheless, the differences due to the choice of reference point are rather small.

Figure 24 draws together the results of the 1–10 kHz measurements. The overall amplitude maximum (LL) has been set to the target values of 75, 65 or 55 dB. It can be seen that in the female subjects the mean dB values differ from each other by more than 10 dB, i.e. the difference between the overall SPLs in the loudness conditions, in the 1–5 kHz area. The male

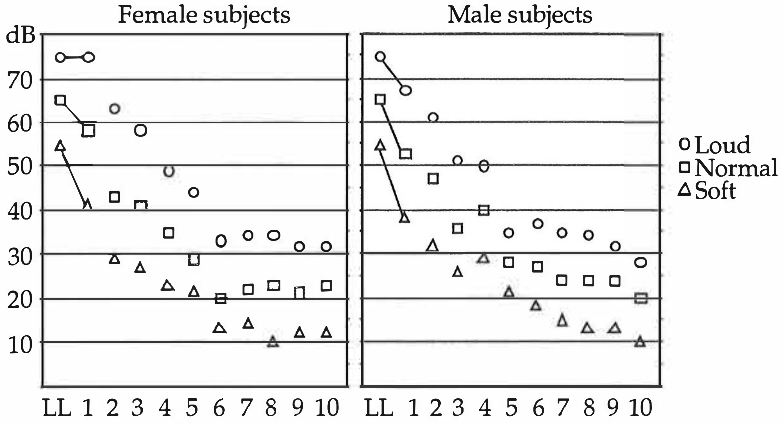


FIGURE 24 Mean dB values in the 1–10 kHz measurements in the three loudness conditions, arranged according to the targeted vocal intensity. The lines connect the maximum dB values (LL, set to 55, 65 and 75 dB) and the dB values at 1 kHz, indicating the spectral slope.

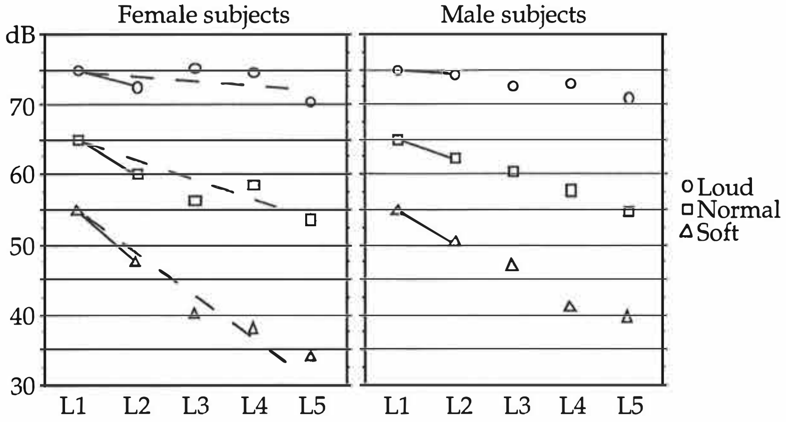


FIGURE 25 Mean dB values in the L2–L5 measurements in the three loudness conditions, arranged according to the targeted vocal intensity. Continuous lines connect L1 (placed at 55, 65 and 75 dB) and L2; in female subjects the broken lines approximate the spectral slope in the L1 to L5 area.

subjects show similar results in the 1–4 kHz area. Mean differences greater than 10 dB illustrate the existence of spectral differences in the three loudness conditions (in addition to the overall SPL differences). In the 6–10 kHz area in the female subjects and in the 5–10 kHz area in the male subjects the means are closer to the overall 10 dB distance, suggesting that these high frequencies do not differentiate loud, normal and soft voices.

To help better appreciate the spectral slope, connecting lines have been drawn in Figure 24 between the point of reference (LL) and the dB value at 1 kHz. The steepness of the slope clearly follows loudness: it is steeper in soft voice and more level in loud voice, especially in the female subjects.

Figure 25 sums up the data for the L2–L5 measurements in the same manner. L1 has been set to the target values of 75, 65 and 55 dB in the three loudness groups.

The measurement points fan out in a similar manner to that in Figure 24. The lowest harmonics clearly differentiate loud and soft voices: in loud voice the level of these harmonics is about the same as L1, whereas in soft voice L1 clearly dominates.

The spectral slope is indicated in Figure 25 by continuous lines connecting L1 and L2. The male subjects clearly show an almost level slope in loud voice and a steeper slope in soft voice. In addition to the line connecting L1 and L2, the general spectral slope in the female subjects is also sketched by broken lines from L1 to L5. These lines shows a similar pattern to those connecting L1 and L2 in male subjects, although the slopes are steeper.

5.3 Support in singing voice

Support and lack of support in singing was mainly investigated by acoustic methods in the present study. Supported and unsupported singing by eight Finnish singers was analyzed by means of average spectrum analysis. However, the results of the acoustic measurements are also related to perceptual ratings of supported and unsupported singing.

5.3.1 Acoustic measurements

Sequences of [pɑ]-syllables were sustained by Finnish singers at two pitches, low and high. The subjects were instructed to sing the low notes below their subjective register transition area and the high notes above the transition area. The average fundamental frequency of the singers in the various tasks has been described in Table 5 on page 57. The low notes were usually sung at about 200 Hz (from F3 to A#3), even though one subject (subject 6) sang much higher (G#4). There was more variation in the pitch of the high notes; their range was from about 340 to 770 Hz (F4 to G5).

The relative levels of the first and second harmonic as well as the "low" and "high" resonances (0.5 to 2 kHz and above 2 kHz, respectively) were measured with the highest amplitude level as the point of reference. The abbreviations L1, L2, LowR and HighR are used. In the high notes LowR usually equals L3 (the level of the third harmonic) and HighR equals L4 (the level of the fourth harmonic).

The results of L1, L2, LowR and HighR measurements for all singers with low and high notes pooled and for low and high notes separately are given in Figure 26 (overleaf) by means of means and standard deviations. The results for all subjects form a baseline for comparisons between various groups.

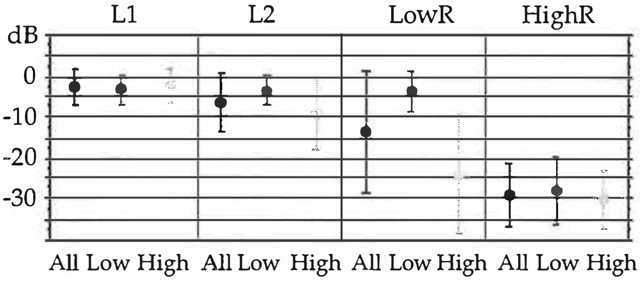


FIGURE 26 Levels of L1, L2, LowR and HighR in all subjects (All) and in low and high notes (Low and High); means and standard deviation are indicated by dots and whiskers.

Low and high differ from each other in all measures. In the L2, LowR and HighR measurements the mean level is lower in the higher notes than in the low notes. In the L1 measurements the mean level is higher in the high notes than in the low notes. The differences between the means are statistically significant (L1: F = 10.3**, L2: F = 59.8***, LowR: F = 265.8***, HighR: F = 6.4*). The means at LowR differ by as much as -4 dB (low notes) and -24 dB (high notes).

Probably, such large differences are partly due to vocal tract resonances. If the fundamental frequency is about 200 Hz, then L2 is measured at about 400 Hz and LowR somewhere in the F1–F2 region (typically between 600 and 1200 Hz), resulting in high amplitude. However, if F0 is about 600 or 700 Hz, then L2 is measured at 1200–1400 Hz and LowR at 1800–2100 Hz (outside the F1–F2 region), resulting in low amplitude. On the other hand, there are probably also phonatory differences between low and high notes, as the L2 values in high notes are not higher than in low notes, even though L2 is located in the F1 area in the high notes (which should increase amplitude), but below the lowest resonance in the low notes (which should decrease amplitude).

Be that as it may, it is clear that how harmonics are subjected to vocal tract resonances depends on fundamental frequency. In an attempt to reduce such variation, low and high notes are treated separately in the subsequent analyses.

Figure 27 compares supported and unsupported singing separately in low and high notes. In the low notes, L2 and HighR are on the average lower in unsupported voice than in supported voice. The differences are statistically significant (L2: F = 9.0**, HighR: F = 4.2*). In the high notes, HighR behaves as in the low notes (HighR: F = 10.7**). Thus, both low and high supported voices have more amplitude above 2 kHz than unsupported voices.

In the high notes, L1 has a higher level in unsupported voice than in supported voice; the difference between means is significant (L1: F = 9.7**). In fact, in L1 the mean dB value for unsupported voice is rather close to zero. (If the mean L1 level were zero, this would indicate spectral domination by the first harmonic in all cases.) Thus, in the high notes the first harmonic more often dominates the spectrum in unsupported voice

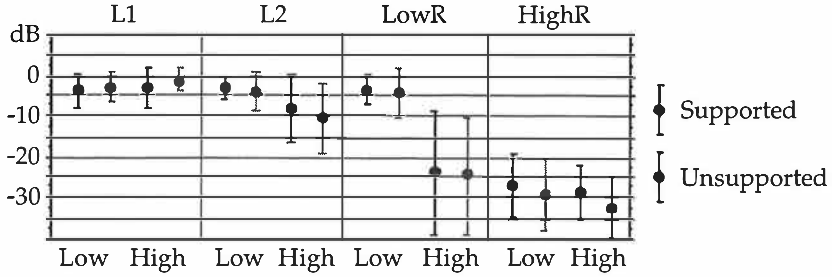


FIGURE 27 Levels of L1, L2, LowR and HighR in supported and unsupported vowels (Sup and Un) separately for low and high notes (Low and High). Means and standard deviation are indicated by dots and whiskers.

than in supported voice (-1 dB vs. -3 dB). In the low notes the direction is the same, even though the difference is not significant.

The comparisons presented above are complemented by computing the relation of the level of the second harmonic (L2) to that of the fundamental (L1). If L2 is higher than L1, the difference in level is positive; if L1 is higher than L2, the difference in level is negative. The results are given in Table 10 together with statistical comparisons by means of analysis of variance.

TABLE 10 Comparison of L1-L2 differences in dB of all subjects in all conditions (All), in low vs. high notes (Low, High) and in supported and unsupported voice (Sup, Un), separately for low and high notes. Means, standard deviation and statistical significance by ANOVA are shown.

Group & Task	x	s	Significance
All	-3.8	9.5	
Low	-0.2	5.0	F = 54.6, p = .0001***
High	-7.4	11.3	
Low/Sup	0.9	4.9	F = 8.0, p = .005**
Low/Un	-1.2	4.8	
High/Sup	-5.0	11.8	F = 7.2, p = .008**
High/Un	-9.8	10.4	

The L1-L2 measure discloses systematic differences between the groups. It is positive in the low supported voices, indicating that the level of the second harmonic is higher than that of the first harmonic. It is more negative in the high notes than low notes and more negative in unsupported voice than supported voice. In other words, the L1-L2 difference is larger in unsupported voice than in supported voice, indicating a steeper slope in unsupported voice. The differences between the means are statistically significant (see Table 10). In fact, part of this information can be extracted from Figure 27 by comparing the levels of L1 and L2; however, it is more easily derived from the mean L1-L2 differences presented in Table 10. Some trends start to emerge, but a more detailed analysis is needed to establish whether part of the variation can be explained by other factors, such as the position of the analyzed vowels in the syllable sequence (initial vs. final) and gender.

Two sets of vowel samples were chosen for spectrum analysis, both at the beginning and towards the end of each syllable sequence. A comparison of the sustained vowels on the basis of their position in the syllable sequence (initial vs. final) was carried out. The means of L1, L2, LowR and HighR as well as of L1-L2 differences for initial and final vowels, carried out separately for low and high notes, show differences of the order of 0 to 1 dB only; none of the differences are statistically significant. Consequently, initial and final segments have been pooled in comparisons of voice type and gender. Below, the measurements for female vs. male singers and for each singer will be examined in more detail.

The results of the L1, L2, LowR and HighR measurements for the female and male subjects are shown in Figure 28. Supported and unsupported voices are pooled.

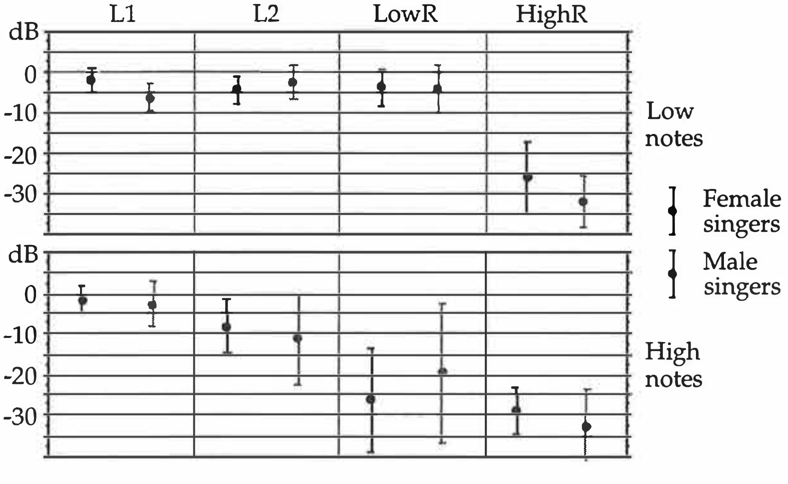


FIGURE 28 Comparison of the levels of L1, L2, LowR and HighR for female and male singers in low notes (upper panel) and high notes (lower panel); means and standard deviation are indicated by dots and whiskers.

Female and male singers show differences in most measures. In the low notes (upper panel), the mean is higher in the female subjects than male subjects in the L1 and HighR measures. In L2 the male subjects have a higher mean dB value. The differences are statistically significant (L1: F = 59.2***, L2: F = 11.2***, HighR: F = 18.4***). In the LowR measures the difference is small.

In the high notes (lower panel) the mean dB values for the men are generally lower than those for the women, but at LowR the men have higher values. All differences between the means are statistically significant (L1: F = 4.5*, L2: F = 5.5*, LowR: F = 8.6**, HighR: F = 6.9**). The high mean amplitude in the men at LowR may be due to vocal tract resonance: LowR measurements have been taken at about 1 kHz, i.e. a typical F1 and F2 area for the [ɑ]-vowel in male voices, whereas in female voices, due to their higher fundamental frequency, the area of measurement is higher, about 1.5–2 kHz, where [ɑ]-vowels usually do not have resonances.

Figure 28 also shows that L1 is in general higher than L2, but in low notes the male voices have higher L2 than L1. The difference in the L1-L2 measure is significant in the low notes: mean L1-L2 is -2.5 dB in the female voices and 3.7 in the male voices (F = 90.2***). The L1-L2 difference in the high notes is not significant. The tentative conclusion is that fundamental frequency dominates in the spectra of low notes in female voices but not in male voices. Further gender differences will be presented below in individual comparisons of supported and unsupported voice.

To return to support in singing, Figure 29 (overleaf) shows the mean values for L1, L2, LowR and HighR in supported and unsupported voice, separately for female and male singers and for low and high notes.

On the whole, the differences between supported and unsupported voice are rather small. In the low notes, the female subjects have lower mean levels in L1 and L2 in unsupported voice than in supported voice; the differences are significant (L1: F = 4.8*. L2: F = 10.7**). The other measures in the female singers show very little difference. The L1 and L2 measurements will be examined more closely below.

In the male singers, the mean level of L1 in the low notes is significantly higher in unsupported than in supported singing (F = 15.0***). On the other hand, the level of HighR is lower in unsupported than in supported singing (F = 29.3***). Thus, unsupported singing of low notes in male singers is characterized by a lower amplitude in the higher frequencies and higher amplitude in the fundamental frequency. The other measures among the male singers differ little.

In the high notes, the differences in the mean levels of the female singers singing in supported and unsupported voice are small and there is much variation. On the other hand, the male singers show significant differences between supported and unsupported voice. The dB levels of L2, LowR and HighR are on the average higher in supported voice than unsupported voice. However, at L1 the mean dB value is lower in supported voice than in unsupported voice. The absence of any standard deviation around the mean of L1 for unsupported voice indicates that in the male subjects the level of this harmonic was without exception the highest in the spectra for unsupported vowels. The differences among the male singers are significant (L1: F = 23.2***, L2: F = 5.3*, LowR: F = 4.1*, HighR: F = 29.3***).

In sum, there are distinct gender-related differences in the L1, L2, LowR and HighR measures. Among the female singers the measures used fail for the most part to distinguish between supported and unsupported voice. Among the male singers, on the other hand, the measures distinguish between supported and unsupported voice rather well. It is clear that the lack of significant differences between means may be due to large variation; indeed, there is much variation in some measures. Therefore, the measurements will also be reported for each individual.

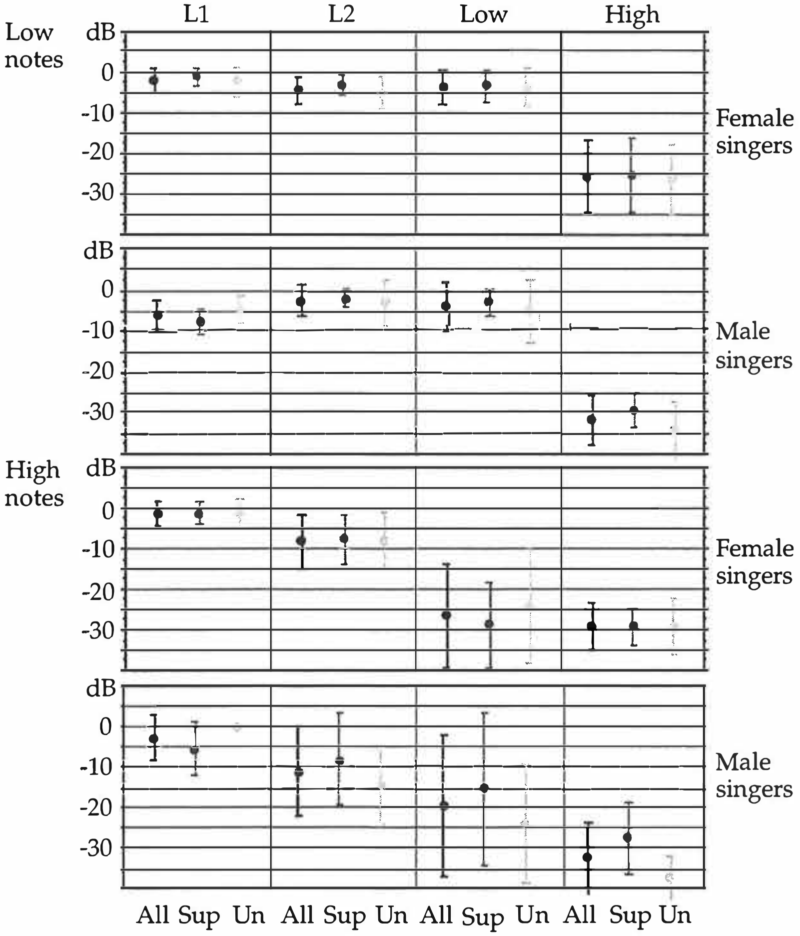


FIGURE 29 Comparison of the levels (means and standard deviation, indicated by dots and whiskers) of L1, L2, LowR and HighR in supported (Sup) and unsupported (Un) voice, separately for low and high notes and for female and male singers. The mean level (All) for both supported and unsupported voice is shown for comparison.

As in Table 10 above (on page 85), the L1-L2 measure can be used to describe differences between supported and unsupported voice and between female and male voice. Table 11 gives the mean L1-L2 differences together with the standard deviation and the result of significance testing by analysis of variance. The table starts by comparing the L1-L2 differences across female and male subjects. Female and male productions of low and high notes are then contrasted. Finally, supported and unsupported singing is compared among female and male singers in low and high notes.

Table 11 shows that the mean L2 value in relation to L1 is lower in the female subjects than male subjects (-4.6 dB vs. -2.3 dB). A closer look at the low notes reveals an accentuated relation: male L1-L2 values are positive (3.7 dB) and female values negative (-2.5 dB). However, a comparison

TABLE 11 Comparison of L1-L2 difference in dB of all subjects in both low and high notes (All), in low vs. high notes (Low, High) and in supported and unsupported voice (Sup, Un) separately for female and male subjects. Means, standard deviation and statistical significance by ANOVA are shown.

Group & Task	x	s	Significance
All	-3.8	9.5	
Female	-4.6	6.9	F = 4.5, p = .03*
Male	-2.3	12.6	
Female/Low	-2.5	2.2	F = 90.2, p = .0001***
Male/Low	3.7	5.8	
Female/High	-6.8	9.0	F = .7, p = .40
Male/High	-8.4	14.5	
Female/Low/Sup	-2.1	2.3	F = 3.3, p = .07
Female/Low/Un	-2.9	2.1	
Female/High/Sup	-6.6	8.0	F = .05, p = .82
Female/High/Un	-7.0	9.9	
Male/Low/Sup	5.9	11.8	F = 7.2, p = .008**
Male/Low/Un	-9.8	10.4	
Male/High/Sup	-2.4	16.1	F = 12.1, p = .001**
Male/High/Un	-14.3	9.7	

of female and male L1-L2 values in the high notes does not yield a significant difference. In the female voices, the mean L1-L2 values between supported and unsupported singing show little difference; the differences are not significant either in respect of low or high notes. On the other hand, the male voices show a significant difference between supported and unsupported singing: in the low notes the mean of the L1-L2 difference is 5.9 dB in supported voice and -9.8 in unsupported voice and in the high notes -2.4 dB in supported voice and -14.3 dB in unsupported voice.

Thus, the L1-L2 measures are in line with the L1, L2, LowR and HighR measurements reported in Figure 28. The female and male subjects differ in the measurements when supported and unsupported voices are pooled. When examining supported vs. unsupported singing, the male subjects show significant differences in almost all measurements and the female subjects in a few measures only. The measurements reveal a very consistent difference in supported and unsupported singing among the men, but little difference among the women.

Above, the groups have been compared. Below, the results of the measurements for each individual will be presented. As before, the results of the L1, L2, LowR and HighR measurements will be reproduced separately for the low and high notes. Figure 30 (overleaf) shows the results for the low notes and Figure 31 (on page 91) for the high notes. Singers 1–3 are male and singers 4–8 female. The patterns in Figures 30 and 31 show considerable variation between the singers. In several singers the dominating peak in the

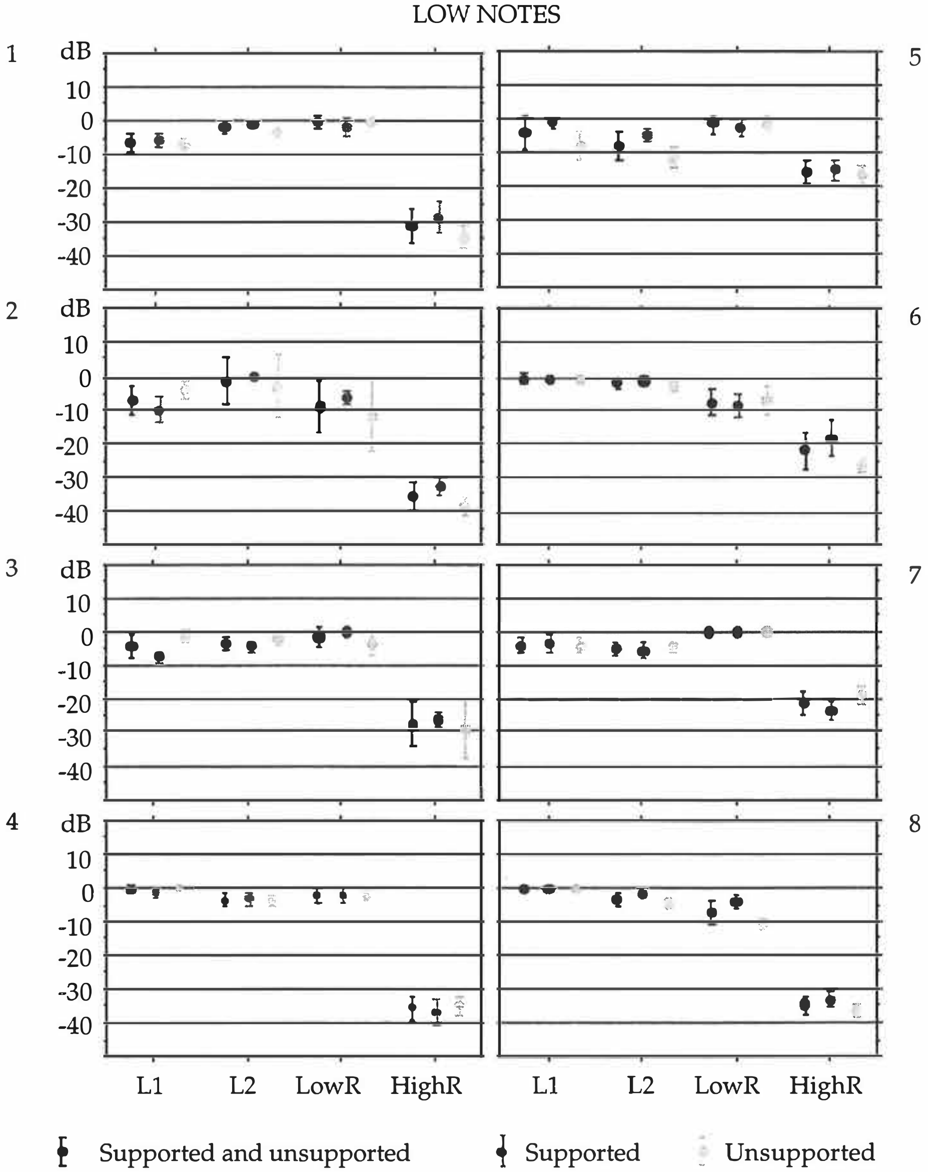


FIGURE 30 Comparison of the levels of L1, L2, LowR and HighR in low notes for each subject (1–8, of which 1–3 male and 4–8 female). Means and standard deviation (in dots and whiskers) are given in triplets, where the value for both supported and unsupported voice is on the left, for supported voice in the middle and for unsupported voice on the right.

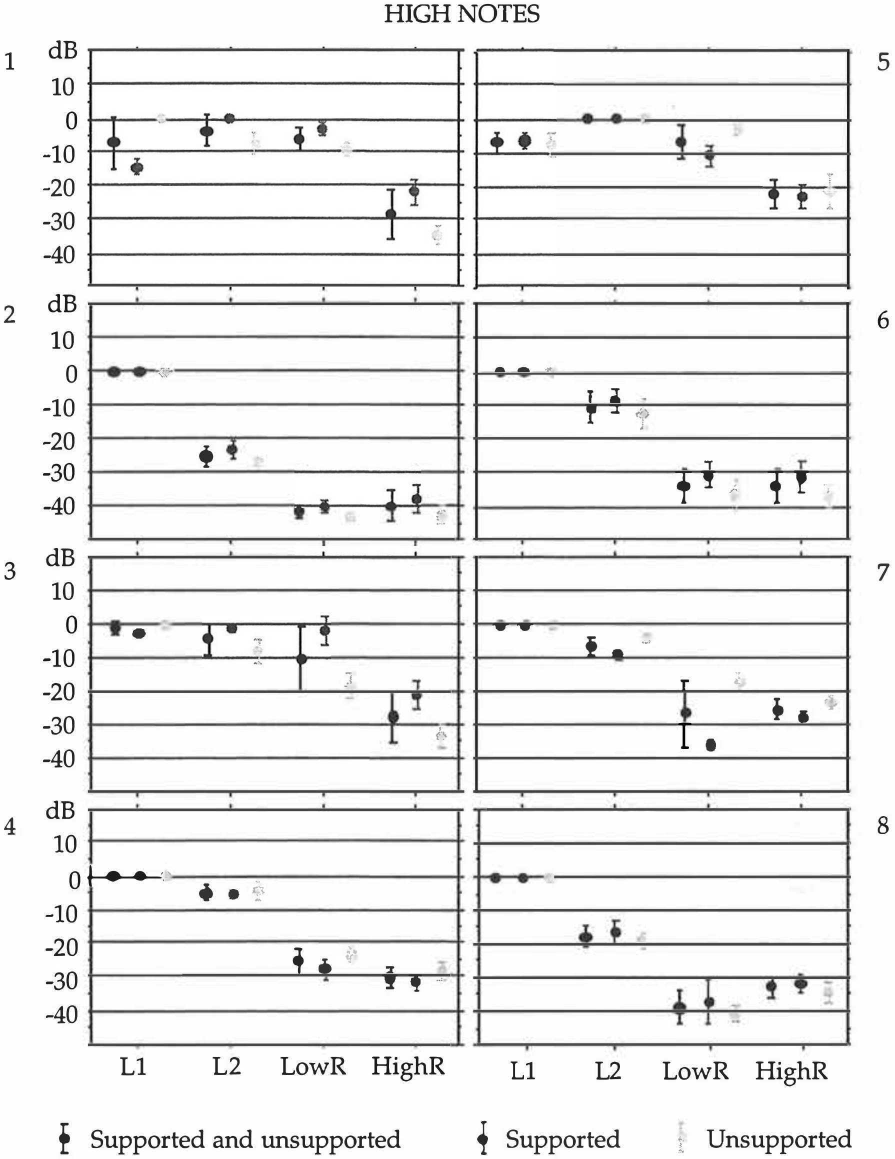


FIGURE 31 Comparison of the levels of L1, L2, LowR and HighR in high notes for each subject (1–8, of which 1–3 male and 4–8 female). Means and standard deviation (in dots and whiskers) are given in triplets, where the value for both supported and unsupported voice is on the left, for supported voice in the middle and for unsupported voice on the right.

TABLE 12 L1-L2 differences in dB in each subject in all notes (All) and in supported and unsupported voice (Sup, Un), separately for low and high notes (Low, High); means, standard deviation and statistical significance by ANOVA of the difference between supported and unsupported voice is shown.

Note	Singer	Task	x	s	Significance
Low	1	All	4.4	1.5	
		– Sup	4.7	1.6	F = .7, p = .43ns
		– Un	4.1	1.5	
	2	All	5.6	9.1	
		– Sup	10.0	3.8	F = 5.8, p = .03*
		– Un	1.2	10.9	
	3	All	1.1	3.1	
		– Sup	3.1	1.1	F = 16.5, p = .0007***
		– Un	–1.0	3.0	
	4	All	–2.9	2.6	
		– Sup	–2.2	3.4	F = 1.3, p = .27ns
		– Un	–3.6	1.6	
	5	All	–3.8	1.7	
		– Sup	–4.2	1.0	F = 2.7, p = .12ns
		– Un	–3.4	2.1	
	6	All	–1.2	2.0	
		– Sup	–0.6	2.1	F = 1.0, p = .33ns
		– Un	–1.9	1.7	
	7	All	–1.2	1.3	
		– Sup	–1.9	1.2	F = 6.9, p = .02*
		– Un	–0.6	1.0	
	8	All	–3.2	2.0	
		– Sup	–1.6	1.4	F = 41.6, p = .0001***
		– Un	–4.8	0.8	
High	1	All	3.5	11.6	
		– Sup	14.5	2.3	F = 356.7, p = .0001***
		– Un	–7.5	2.9	
	2	All	–25.3	2.9	
		– Sup	–23.3	2.8	F = 15.9, p = .0009***
		– Un	–27.2	1.3	
	3	All	–3.3	5.7	
		– Sup	1.6	1.4	F = 63.3, p = .0001***
		– Un	–8.2	3.6	
	4	All	–5.0	2.3	
		– Sup	–5.3	1.3	F = .3, p = .57ns
		– Un	–4.6	2.5	
	5	All	7.1	2.4	
		– Sup	6.7	1.8	F = .7, p = .42ns
		– Un	7.6	3.0	
	6	All	–11.6	5.5	
		– Sup	–8.7	3.5	F = 7.7, p = .012*
		– Un	–14.5	5.7	
	7	All	–6.7	2.8	
		– Sup	–9.1	1.5	F = 55.6, p = .0001***
		– Un	–4.3	1.4	
	8	All	–17.8	3.2	
		– Sup	–16.5	3.6	F = 3.6, p = .07ns
		– Un	–19.1	2.2	

spectrum is L2 or LowR (subjects 1, 2, 3, 5 and 7 in the low notes and subjects 1 and 5 in the high notes). The other subjects show L1 dominance.

Before further analyzing the direction of differences in level between supported and unsupported voice in each singer, the L1, L2, LowR and HighR measurements given are complemented by L1-L2 measures. The mean L1-L2 difference for each singer together with the standard deviation is given in Table 12 (on page 92); in the upper part for the low notes and in the lower part for the high notes. The mean for all low or high notes is shown first, then the means for supported and unsupported voice. The table also shows the results of analysis of variance used to test the significance of the differences in the means between supported and unsupported voice.

The results presented in Table 12 on supported and unsupported voice tie in with those presented in Figures 30 and 31. The figures demonstrate that there are differences in the mean dB levels in both directions: a given level in one singer may be higher in supported voice, in another in unsupported voice. The direction of the differences is further examined in Table 13 (overleaf).

Some singers show several significant differences between supported and unsupported voice in the L1, L2, LowR and HighR measures. On the other hand, there are also singers who have no or few significant differences: singer 4 in the low notes and singer 8 in the high notes show no significant differences and singers 6 and 7 in low notes and singers 5 and 8 in high notes show few differences only.

Table 12 shows that the differences of the L1-L2 means for supported and unsupported voice are statistically significant in about half of the singers. As in Figures 30 and 31 there are differences in the mean dB levels in both directions: in one singer the level of L2 may be higher than that of L1 in supported voice, in another in unsupported voice. The direction of these differences too is examined in Table 13, which summarizes the differences between supported and unsupported voice presented in Figures 30 and 31 and in Table 12 and complements Figures 30 and 31 with information about statistical significance.

In Table 13, ▲ indicates that a measure (L1, L2, LowR, HighR or L1-L2) is higher in supported than unsupported voice; ▼ indicates that a measure is lower in supported voice than in unsupported voice. Statistical significance is indicated by asterisks. The signs that lack asterisks indicate the direction of the differences in amplitude level that exceed the arbitrary threshold of 2 dB. Such differences are not statistically significant; they are included to help trends emerge in the data.

In the male subjects supported voices rather consistently have a higher dB level than unsupported voices in all measures except L1. But a low level of L1 actually points to the same conclusion as a high level in the other measures: the male subjects differentiate supported and unsupported voice by means of a higher level in the harmonics above the fundamental frequency in supported voice, i.e. a more level spectral slope in supported voice and a steeper slope in unsupported voice.

TABLE 13 Direction of difference in level differences at L1-L2, L1, L2, LowR, HighR and in eight singers singing low and high notes in supported and unsupported voice. ▲ indicates that the mean level is higher in supported voice than in unsupported voice and ▼ indicates the opposite (see text).

Low notes		L1-L2	L1	L2	LowR	HighR
Male Ss	1	–	▲*	▲**	–	▲**
	2	▲*	▼**	▲	▲	▲***
	3	▲***	▼***	▼*	▲**	▲
Female Ss	4	–	–	–	–	▼
	5	–	▲***	▲***	–	▲
	6	–	–	–	▼	▲***
	7	▼*	–	–	–	▼**
	8	▲***	–	▲***	▲***	▲**

High notes		L1-L2	L1	L2	LowR	HighR
Male Ss	1	▲***	▼***	▲**	▲***	▲***
	2	▲***	–	▲***	▲***	▲***
	3	▲***	▼***	▲***	▲***	▲***
Female Ss	4	–	–	–	▼**	▼**
	5	–	–	–	▼***	–
	6	▲*	–	▲*	▲**	▲**
	7	▼***	–	▼***	▼***	▼***
	8	▲	–	▲	▲	▲

Female singers do not show unified behavior: two groups can be discerned. Two of the singers show a pattern similar to that of the male singers: singers 6 and 8 generally have a higher level in supported voice than in unsupported voice. The direction of the differences in level is very clear in the low notes for singer 8 and in the high notes for singer 6. In the low notes of singer 6 the measure HighR is higher in supported voice than in unsupported voice. In the high notes singer 8 shows a consistently higher level in supported than unsupported voice (even though the difference is not statistically significant).

The other female group consisting of singers 4 and 7 is characterized by frequent ▼'s or by no significant or even systematic differences. This group often shows lower levels in supported voice than in unsupported voice, i.e. a steeper or a more negative spectral slope in supported voice and a more level slope in unsupported voice. There are no consistent differences at L1, as L1 is often the highest peak in female singers.

Singer 5 appears to be an exception. In low notes, she clearly has a higher level in supported voice than in unsupported voice at L1 and L2. However, as the highest peak here is at LowR (see Figure 30 on page 90), the high levels in supported voice at L1 and L2 indicate a less positive slope in supported voice than in unsupported voice. Thus, singer 5 resembles the second group, consisting of singers 4 and 7, in showing a more negative slope in supported voice than in unsupported voice.

Problems pertaining to the interpretation of the mean amplitude levels at the measurement points will be discussed in chapter 6. The results presented above will be summarized in chapter 5.3.3 after relating the acoustic measures to perceptual ratings of supported and unsupported voices.

5.3.2 Acoustic measurements and perceptual evaluation

Listening tests have been administered by Aatto Sonninen and the present writer to evaluate the supported and unsupported voices on two visual analog scales: supported vs. unsupported singing and good vs. poor voice quality. Analyses have shown a strong positive correlation between the two scales. The perceptual ratings reported below have been obtained with the good vs. poor quality scale, where 100 indicates maximally good and 0 maximally poor.

In this chapter, perceptual ratings are related to acoustic observations. First, the quality ratings for several groups and also for each singer are described and then the relationship between acoustic and perceptual measures is examined. The mean quality rating for all singers singing low and high notes in supported and unsupported voice is 35 (on the scale 100–0). There is a lot of variance, but it will be omitted here; the mean rating of all raters for each sample is taken as the basis for the description. This is a source of error, but the aim here is to examine the relation between overall quality and acoustic measures, for which purpose the mean rating of each sample is sufficient.

In analysis of variance, the means for the perceptual ratings showed a statistically significant difference between low and high notes (29 and 40, respectively; F = 8.1**) and between supported and unsupported voice (40 and 30, respectively; F = 7.3**). The difference between initial and final segments is not significant (36 and 34, respectively; F = 0.4ns). Female singers were assessed better than male singers (39 vs. 28; F = 6.1**).

Table 14 (overleaf) displays a comparison of the mean perceptual ratings of supported and unsupported voices separately in female and male singers and in low and high notes.

TABLE 14 Means of perceptual ratings of supported and unsupported voices in female and male singers and in low and high notes.

	Female+Low	Female+High	Male+Low	Male+High
supported	29	60	37	30
unsupported	25	45	30	17

Supported voices receive a higher rating in all groups. The difference between supported and unsupported voice is greater in high notes than in low notes. The voices of female and male singers show diametrically opposed behavior: female voices receive higher ratings in high notes than in low notes, whereas male voices receive higher ratings in low notes than in high notes. Thus, trends appear to emerge.

Table 15 presents the mean values attributed to each singer in low and high notes both in supported and unsupported voice. The mean for all samples is also shown for each singer.

TABLE 15 Means of perceptual ratings of each singer in all samples (All), in low and high notes (Low, High) and in supported and unsupported voice (Sup, Un).

Singer	All	Low	Low		High	High	
			Sup	Un		Sup	Un
1	27	32	27	37	22	32	12
2	33	38	46	31	27	31	24
3	26	30	38	24	21	27	15
4	30	8	8	9	52	61	44
5	31	22	32	12	40	55	24
6	50	44	39	48	57	63	52
7	43	38	35	41	49	60	38
8	39	23	30	16	55	60	51

There is much variation between the singers. Overall quality varies from 50 (singer 6) to 26 (singer 3). The highest value (63) is for singer 6 in supported high notes and the lowest (8) for singer 4 in supported low notes. In low notes, the supported samples of four singers (2, 3, 5, 8) are rated better than the unsupported samples. On the other hand, the unsupported samples of four singers (1, 4, 6, 7) are rated better than the supported samples. Thus, in low notes both the supported samples and the unsupported samples receive high and low ratings. In high notes, on the other hand, the supported samples of all singers are rated better than the unsupported samples.

The male singers (1–3) do not manifest large differences in the ratings of low and high notes, even though the ratings are lower in high voice. The female singers, on the other hand, cluster in two groups: one group consists of those whose ratings differ rather little between low and high notes, even though the high notes receive somewhat higher ratings. This group consists of singers 6 and 7. The other group is characterized by rather low ratings in low notes and much higher ratings in high notes. This group comprises singers 4, 5 and 8.

These perceptual results have been related to the results of the acoustic measurements by calculating a Spearman rank correlation between the mean perceptual rating for each sample and the acoustic measures L1, L2, LowR, HighR and L1-L2. The results are given for all subjects, for female and male singers separately, for low and high notes, and supported and unsupported voice. In addition, the results are given for female singers with supported voice and with unsupported voice as well as for male singers with supported voice and with unsupported voice. The correlation coefficients which have been corrected for ties are shown in Table 16.

TABLE 16 Spearman rank correlation (r) and statistical significance (p) between perceptual ratings and acoustical measures (L1-L2, L1, L2, LowR, HighR) for all samples, in low and high notes, for female and male samples, in supported and unsupported voice and separately in supported and unsupported voice for female singers (Female/Sup, Female/Un) and male singers (Male/Sup, Male/Un).

	L1-L2		L1		L2		LowR		HighR	
	r	p	r	p	r	p	r	p	r	p
All	-.09	.47ns	.11	.41ns	-.06	.66ns	-.38	.002**	.14	.27ns
Female	-.42	.009**	.29	.07ns	-.32	.04*	-.69	.0001***	-.05	.78ns
Male	.54	.01*	-.63	.003**	.50	.02*	.34	.10ns	.16	.45ns
High	-.08	.67ns	.19	.30ns	-.07	.70ns	-.33	.07ns	.06	.74ns
Low	.58	.001**	-.26	.15ns	.41	.02*	.03	.89ns	.28	.12ns
Supported	-.23	.19ns	.26	.14ns	-.21	.23ns	-.55	.002**	-.02	.93ns
Unsupported	-.04	.83ns	.05	.80ns	.03	.88ns	-.30	.09ns	.03	.87ns
Female/Sup	-.45	.047*	.32	.16ns	-.43	.06ns	-.75	.001**	.05	.85ns
Female/Un	-.40	.08ns	.29	.20ns	-.27	.23ns	-.65	.004**	-.24	.31ns
Male/Sup	.23	.44ns	-.29	.34ns	.29	.34ns	-.12	.69ns	-.17	.58ns
Male/Un	.59	.05ns	-.74	.014*	.55	.07ns	.49	.10ns	-.22	.47ns

The perceptual ratings in all samples show a significant correlation with LowR. The correlation is negative, i.e. high quality ratings and low values for LowR go together – or vice versa, low quality ratings and high values for LowR. In other words, samples rated high in quality in general have a low value for LowR (level at .5–2 kHz).

The female singers show a statistically significant negative correlation between perceptual ratings and three acoustic measures: L1-L2, L2 and LowR. Consequently, in the female samples, those that have been rated high in quality in general have a lower level of L1-L2, L2 and LowR and vice versa. Male singers show a negative correlation at L1 and a positive correlation in two measures: L1-L2 and L2. Thus, male voices that have been rated high in quality in general have a lower level of L1 and a higher level of L1-L2 and L2.

High notes do not show a statistically significant correlation between any acoustic measures and perceptual ratings, but low notes do: a positive correlation both at L1-L2 and L2, indicating that in low notes good ratings go together with a high level of L1-L2 and L2.

Supported voices show a negative correlation between perceptual ratings and LowR measures, but unsupported voices do not show significant correlations. Thus, good ratings in supported samples go with low levels of LowR.

When supported and unsupported voices are examined separately in the female and male subjects, some statistically significant correlations can be observed: female supported voices show a negative correlation between perceptual ratings and L1-L2 and LowR, and female unsupported voices at LowR, too. The correlations are negative both in supported and unsupported voice, indicating that good ratings go together with a low level of L1-L2 and LowR. In supported male voices none of the correlations between ratings and levels are significant. In unsupported male voices there is a significant negative correlation between perceptual ratings and L1, associating a low level of L1 with high ratings in quality. There is also an almost significant positive correlation between perceptual ratings and L1-L2, suggesting that a high level of L2 in relation to L1 may be associated with high ratings in quality.

Table 17 brings together the observations on correlations between the perceptual ratings and acoustic measures. Where a correlation is positive, there is an association between high perceptual ratings and high values in the acoustic measures. Where a correlation is negative, the association is between high perceptual ratings and low values in the acoustic measures or between low perceptual ratings and high values in the acoustic measures.

TABLE 17 Significant positive and negative correlations between perceptual ratings and acoustic measures (L1-L2, L1, L2, LowR and HighR).

	L1-L2	L1	L2	LowR	HighR
Positive correlation	Low Male Male/Un		Low Male		
Negative correlation	Female Female/Sup	Male Male/Un	Female	All Female Sup	

All the acoustic measures show statistically significant correlations except HighR. There are positive correlations between perceptual ratings and L1-L2 and L2. In low notes, in male singers and in male singers singing in unsupported voice there is a positive correlation between perceptual ratings and L1-L2, and in low notes and in male singers between perceptual ratings and L2. Thus, high levels of L2 in relation to L1 and of L2 by itself are associated with high quality ratings in low notes and in male voices (and also in male unsupported voices).

All samples, female samples and supported voice show a negative correlation between perceptual ratings and LowR: high ratings go with low levels of LowR and low ratings go with high levels of LowR. There are two negative correlations between the ratings and L1-L2, in the female samples and in the female supported samples. Thus, in these groups low levels of L2 in relation to L1 are associated with high quality ratings.

The female and male voices behave in a different way: high L2 levels in relation to L1 are associated with high ratings in the male voices, whereas in the female voices low levels of L2 are associated with high ratings. In other words, in female singers L1 dominance is associated with better perceptual ratings, in male singers better perceptual ratings are associated with lower L1 and higher L2, i.e. dominance of harmonics higher than L1.

These observations point to differential behavior in listeners. When judging female singers, they give better ratings to singing with more L1 dominance. When judging male singers, they give better ratings to singing where the higher harmonics are more dominant.

Supported samples show a negative correlation at LowR: good ratings of supported samples are associated with low levels of LowR. There are further correlations where supported and unsupported samples are involved. In the male unsupported voices there is a positive correlation with perceptual ratings and a high level of L2 in relation to L1. In the female supported samples there is a negative correlation at L1-L2: high ratings are associated with low levels of L2.

Two methodological observations conclude the presentation of the relationship of acoustic measures and perceptual ratings. The first concerns the five acoustic measures used in the correlations. A closer look at the correlations in two measures, L1-L2 and L1, shows that they often are in opposed directions. This is largely a consequence of the measurement procedure (see chapter 4.3.2). If L1 is high (or even the peak with highest amplitude, i.e. the point of reference), L1-L2 is typically low. If L1 is low then L1-L2 is typically high, as L2 (or LowR) must be higher, since otherwise L1 could not be lower. This methodological matter will be brought up for further discussion in chapter 6. However, it has the consequence that Table 17 can be further condensed by omitting L1 and L2 from the results of supported and unsupported voice to be summed up in chapter 5.3.3.

The second methodological observation concerns the correlations between the perceptual ratings and acoustic measures. It would be tempting to compute these correlations for each individual to examine differences between individuals. However, there are not enough data points for this to be feasible.

5.3.3 Synopsis of the results on support

Table 18 combines the statistically significant differences in the mean spectral levels in supported and unsupported voice in three frequency areas: low, i.e. the difference in the levels of the second harmonic and the first harmonic (L1-L2), middle, i.e. the level at the middle frequencies (LowR), and high, i.e. the level at the higher frequencies (HighR). The table shows the groups which have high spectral amplitude (indicated by ▲) and the groups that have low amplitude (indicated by ▼) in a certain frequency area.

On the whole, supported voices have a flatter spectrum than unsupported voices both in low and high notes. The trend shows better in male subjects than female subjects. Both the female and male voices differ statistically significantly in the supported/unsupported comparisons in the L1-L2 measures; only the male voices differ significantly in the LowR and HighR measures. In sum, both female and male voices have a more positive spectral slope in a comparison of the levels of the lowest two harmonics. Only the male subjects have significant differences between supported and unsupported voice in the higher spectral measures – and more significant differences in high notes than low notes.

TABLE 18 Direction of statistically significant differences in spectral slope at three acoustic measures (L1-L2, LowR and HighR) in low and high notes. The groups indicated by ▲ have higher spectral amplitude, those by ▼ lower amplitude. Slashes indicate subgroups, e.g. supported/male indicates a significant difference in the means for supported and unsupported voice in male singers.

Low notes	L1-L2	LowR	HighR
▲	supported male		supported/male female
▼	unsupported female		unsupported/male male
High notes	**L1-L2**	**LowR**	**HighR**
▲	supported male	supported/male male	supported/male female
▼	unsupported female	unsupported/male female	unsupported/male male

In a comparison where support is not taken into account, male subjects generally have a flatter spectrum than female subjects. However, the female subjects show higher mean values at high frequencies (in the HighR measurements).

There is much individual variation in the data. Individual comparisons show that male singers and two of the female singers (6 and 8) tend to differentiate supported and unsupported voice by producing supported voice with a more level spectrum than unsupported voice. Three of the female singers (4, 5 and 7) tend to differentiate supported and unsupported voice by producing supported voice with a steeper spectrum than unsupported voice. In view of such differential behavior in the female singers, it is obvious that the measures employed often fail to show a difference between supported and unsupported voice in the female singers.

Table 19 summarizes the significant correlations between perceptual ratings and main acoustic measures (L1-L2, LowR and HighR).

TABLE 19 Significant positive and negative correlations between perceptual ratings and main acoustic measures (L1-L2, LowR and HighR).

	L1-L2	LowR	HighR
Positive	Low Male Male/Un		
Negative	Female Female/Sup	All Female Sup	

Listeners evaluated female samples and supported samples better if they had a lower level at LowR. A lower level of LowR means a steeper spectral slope in the .5–2 kHz area (in the highest notes somewhat higher, see chapter 4.3.2). This also applies to all samples: a lower level of LowR means a better rating in the samples investigated in this study.

Low notes and male voices were evaluated better if they had a high level in L1-L2, i.e. a relatively level spectral slope. On the other hand, female voices were evaluated better if they had a low level in L1-L2, i.e. a relatively steep spectral slope. Thus, the results are gender-specific: on average, female singers were judged to have better voice if the spectrum was steeper, male singers if the spectrum was shallower.

Low levels in L1-L2 in female supported voices were associated with higher quality ratings. On the other hand, high quality ratings are associated with high levels of L1-L2 in male unsupported voices. As described above, male voices receive high quality ratings with relatively level spectra, female voices with relatively steep spectra with fundamental frequency dominance.

5.4 Covered and open singing voice

5.4.1 Amplitude levels and the voice types

A classically trained female singer sang an ascending series of the vowel [ɑ] in covered voice, in open voice as well as in a loud and open voice type ("young boy's shout"). She sang both forte and piano in covered and open voice. Averaged spectra were computed from the vowels and the dB values at the lowest two harmonics (L1 and L2) and at the LowR and HighR resonances were measured. In addition, the levels of the lowest six harmonics were measured. The dB measurements have been converted to absolute measures as the dB level was measured during the production of each vowel.

The measured dB levels at L1, L2, LowR and HighR were first examined by computing for each measure an average for the whole series of vowels produced at different pitches. Table 20 shows the mean dB levels and standard deviation of L1, L2, LowR and HighR in covered voice (forte and piano), open voice (forte and piano) and in "shouting voice".

TABLE 20 Mean level and standard deviation of L1, L2, LowR and HighR on an absolute dB scale in vowels produced by a female singer at all pitches in the voice types investigated.

		L1	L2	LowR	HighR
Covered, forte	x	86	82	75	58
	s	11.1	9.6	7.1	10.2
Covered, piano	x	74	54	49	35
	s	5.6	5.7	3.9	4.2
Open, forte	x	82	86	93	69
	s	4.8	8.3	6.7	7.9
Open, piano	x	75	74	67	49
	s	14.0	1.1	7.6	4.9
Shout	x	84	97	92	78
	s	14.0	7.5	6.2	7.7

Shouting voice naturally has the highest amplitude (97 dB at L2), then forte singing (93 dB at LowR in forte open, 86 dB in forte covered), then piano singing (75 dB at L1 in piano open and 74 dB in piano covered).

The mean values also show the spectral differences between the voice types. The highest amplitude is at L2 in shouting and at LowR in forte open voice. These two types of singing have the highest dB value above L1. The other voice types, covered forte, covered piano and open piano, share L1 dominance as compared with the other measures. In other words, the voice types systematically arrange themselves on the basis of L1 (as compared with

L2 and LowR): covered voices have a high L1 level, open voices (including shouting as an extreme case) a lower L1 level. Thus, the L1-L2 difference indicates a more positive slope in open than covered voice.

Figure 32 (overleaf) gives a general view of the dB measurements at L1, L2, LowR and HighR in the five voice types, arranged on a logarithmic frequency scale in accordance with the fundamental frequency of the vowel measured. Covered voice is described in the left panels and open voice in the right panels. Shouting voice is included in the right panels as the data points resemble those for open voice. The measurement points are reproduced as such in scattergrams, complemented by a second-order polynomial regression curve to describe the patterning of the data points.

The subject was able to sing a wide range of pitches in covered voice (D3–G#6 or 147–1661 Hz). In open voice the range was much smaller (F#3–A#4 or 185–466 Hz). In piano singing high notes (up to D#6 or 1225 Hz) were possible. In young boy's shout the maximum was lower (D#5 or 622 Hz).

The data points and the polynomials describe the general amplitude level in covered and open voice by means of four measures. However, the differences between covered and open forte voice are not easy to appreciate in Figure 32; they will be examined closer in Figures 33 and 34 below. Piano voice clearly differs from the others by its lower amplitude, although in open piano voice the dB values are usually higher than in covered piano voice. The data points for shouting voice are located in the same area as the data points for forte open voice. The general lines in the over-all amplitude differences are presented in Table 20.

The data points and the polynomials also describe the amplitude levels of the four acoustic measures (L1, L2, LowR and HighR) in relation to the fundamental frequency of the vowels produced in covered and open voice. Three main trends can be seen in the data. Firstly, the curves for piano voice are rather level in both covered and open voice. Secondly, in forte open voice the curves show a steep rise in amplitude in all acoustic measures, but there are no data points above 466 Hz (A#4), as the singer could not sing higher in open voice. Shouting voice in general continues the rising curve of forte open voice. Thirdly, in forte covered voice the curves are in general less steep than in forte open voice and they continue much higher, as the singer was able to produce higher notes in covered voice. To compare covered and open singing, the pitch area common to them (i.e. the pitch range of open singing) was subjected to detailed analysis. First, mean dB values for all pitches are reported, then note by note.

Table 21 (on page 105) gives the mean dB level at L1, L2, LowR and HighR related to the maximum dB level of these measures of each vowel. As these comparisons have been carried out for each vowel and a mean has been computed from the comparisons of each vowel for all vowels produced in a voice type, the mean of the highest peak need not equal zero. The table also gives the standard deviation.

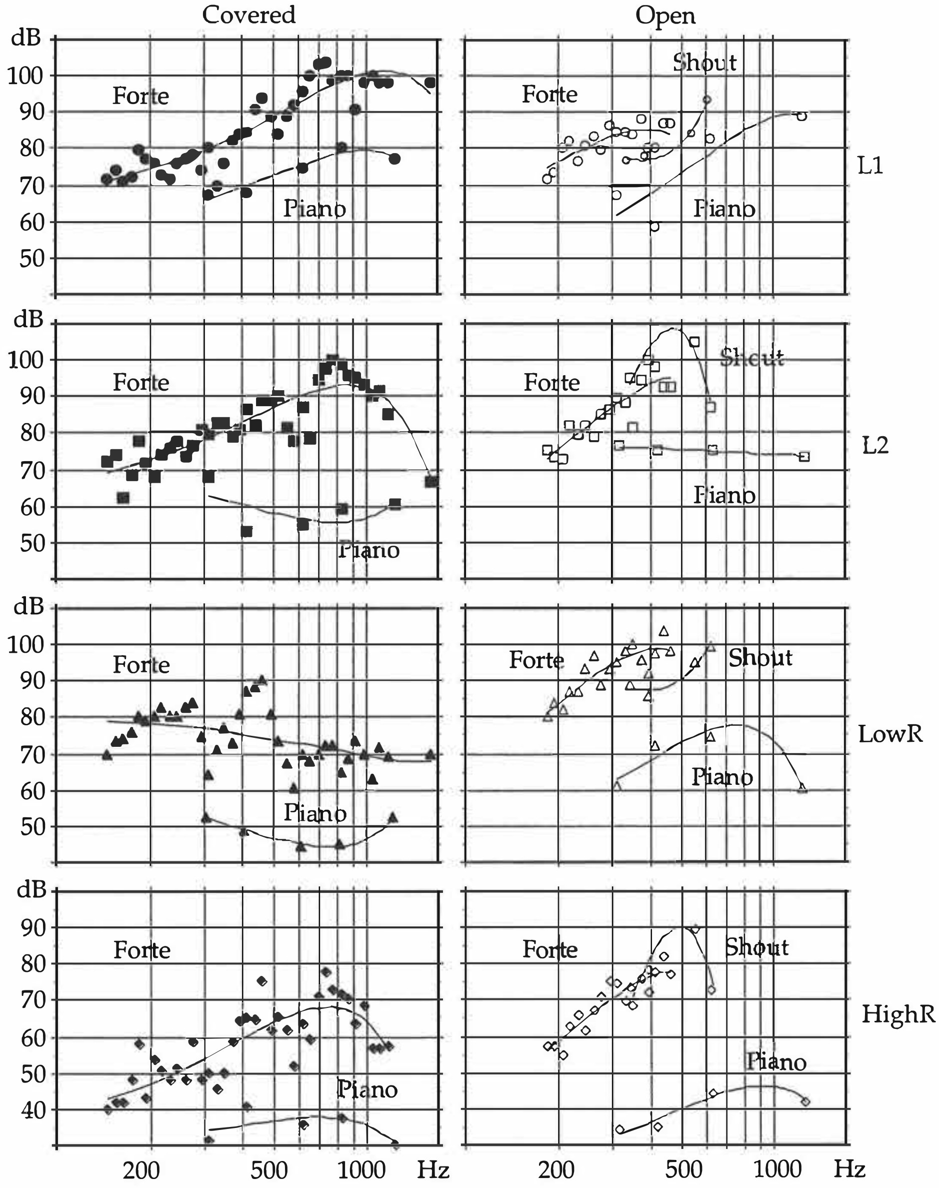


FIGURE 32 The levels of the two lowest harmonics (L1, L2) and the LowR and HighR resonances across the fundamental frequency in covered and open singing (both forte and piano) and in "young boy's shout"; the data points are described by second-order polynomials.

TABLE 21 Mean level (x) and standard deviation (s) of L1, L2, LowR and HighR with maximum level set to zero dB in covered and open forte voice.

		L1	L2	LowR	HighR
Covered, forte	x	–4	–4	–4	–28
	s	4.0	3.9	4.9	5.9
Open, forte	x	–11	–7	–1	–24
	s	4.8	5.4	1.9	4.6

By these measures, which pool vowels produced in the pitch range of F#3–A#4 (185–466 Hz), covered and open voices differ clearly in spectral slope. In covered voice the slope is level. In open voice the slope is rising.

However, such average measures mask pitch-related variation. Figure 33 (overleaf) contrasts the measurement points of Figure 32 across covered and open voice in the pitch area common to both voice types, F#3–A#4 (185–466 Hz). The measured dB values for L1, L2, LowR and HighR have been arranged on a logarithmic frequency scale. The data for covered voice is displayed in the left panels, for open voice in the right panels. The measurement points are reproduced in scattergrams; a second-order polynomial regression curve has been computed to describe the patterning of the data points in each panel.

In general, the levels of L1, L2, LowR and HighR increase with pitch. Another general observation is that the dB levels are higher in open voice than in covered voice. These are expected results. What is more interesting is the shape of the curves; it varies highly systematically across the voice types. In covered voice the shape is convex (with the exception of L2, where it is rather linear). In open voice the shape is concave. In other words, the dB values rise more steeply in open voice than in covered voice. In covered voice the singer "holds back", not letting the amplitude levels rise as fast as in open voice. The mechanism responsible for this will be discussed in chapter 6.

5.4.2 Spectral levels of the lowest two harmonics

By analogy with the analyses of three materials reported in the preceding three sections (5.1 to 5.3) of this study, covered and open voice types have been studied by comparing the level of the second harmonic to that of the first harmonic, i.e. the fundamental. This difference in level (L1-L2) reflects laryngeal behavior in the adduction–abduction continuum, as described in chapter 3.

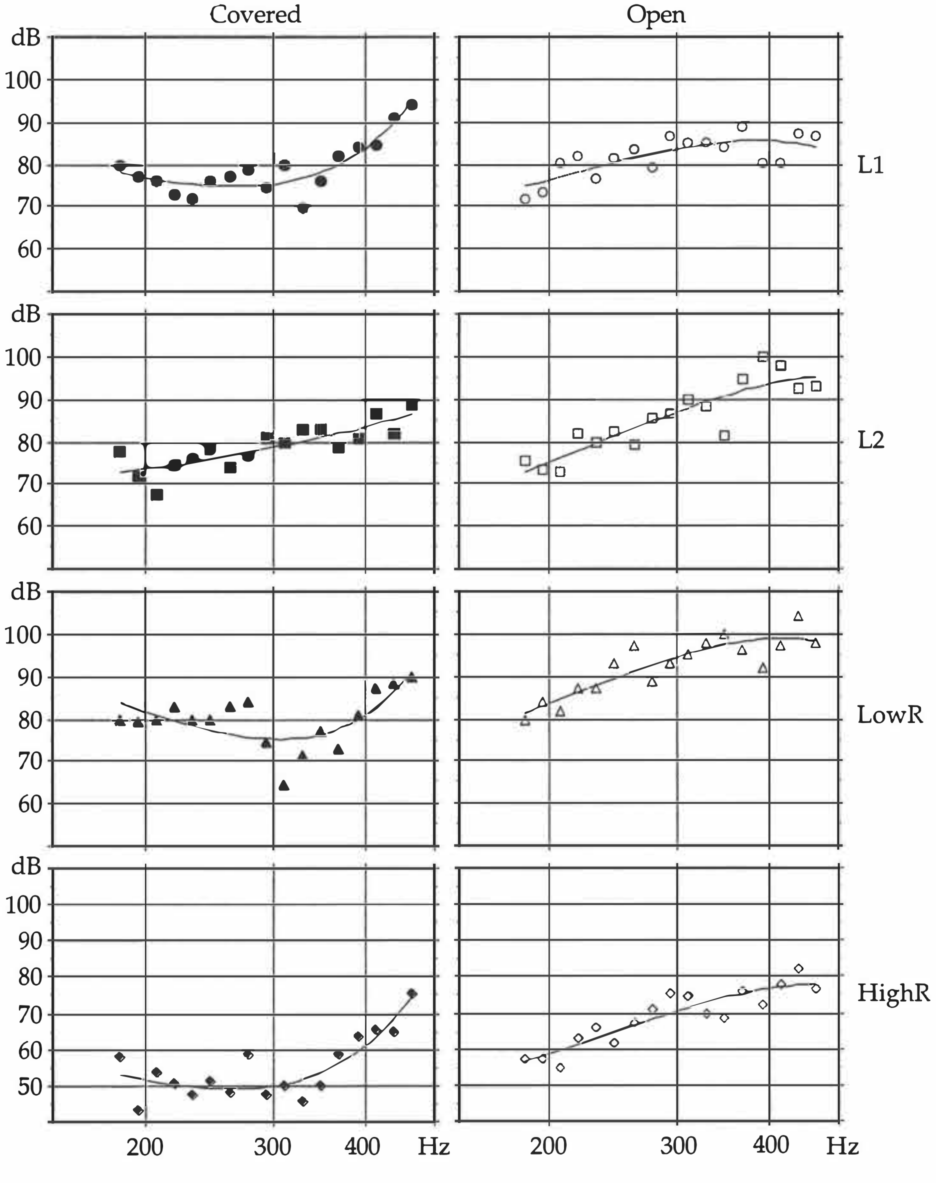


FIGURE 33 Levels of the two lowest harmonics (L1, L2) and the LowR and HighR resonances in covered and open singing (left and right panels) in the pitch area common to both. The horizontal scale gives the fundamental frequency (in Hz) of the vowel measured. The patterning of the data points is described by second-order polynomial regression curves.

Figure 34 shows the relation of the level of the second harmonic to that of the first harmonic in covered voice and in open voice. A positive value indicates that the second harmonic has a higher level than the first harmonic. A negative value indicates that the second harmonic has a lower level than the first harmonic. The L1-L2 level differences are given as data points in scattergrams, patterned by second level polynomial regression curves. The horizontal axis describes pitch as previously.

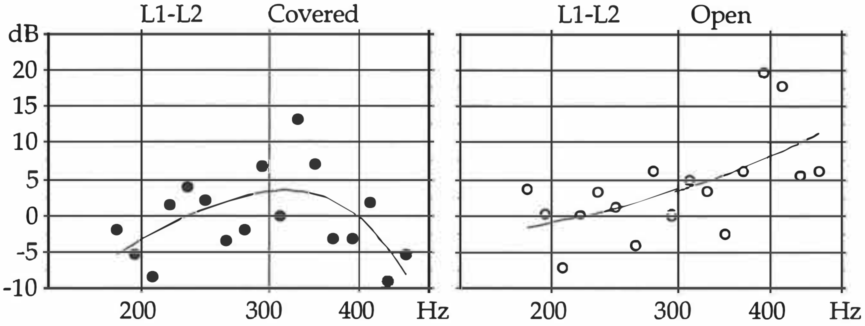


FIGURE 34 L1-L2 difference in covered and open singing in a pitch area common to both. The horizontal scale gives the fundamental frequency (in Hz) of the vowel measured. The patterning of the data points is described by second-order polynomial regression curves.

The second-order polynomials describe the data points well (r = .61* in covered voice and r = .60* in open voice). In open voice the L1-L2 difference increases with fundamental frequency, as indicated by the regression curve. Thus, the second harmonic gains in amplitude over the first harmonic. The same behavior can be seen in covered voice up to about 350 Hz, but then the difference decreases, as indicated by the falling regression curve. The decreasing L1-L2 difference suggests an increasing dominance of L1 above the register transition area. The singer reported her register transition to typically occur in this area.

5.4.3 Harmonics with maximum levels

The amplitude levels of the lowest six harmonics were also measured. The acoustic properties of covered and open voice can be further examined by depicting the maximum dB values for both voice types across pitch and indicating for each data point the harmonic which gives the value (i.e. the maximum peak in the spectrum). This has been done in Figure 35 (overleaf).

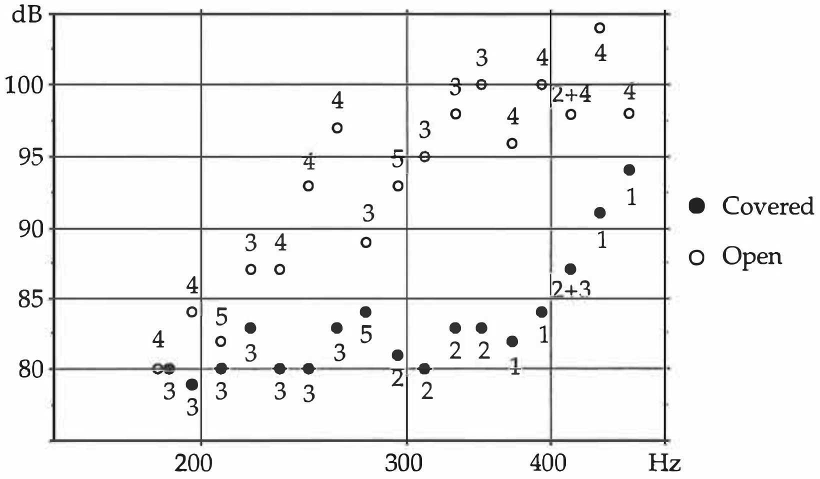


FIGURE 35 The maximum dB values across pitch in covered and open voice together with information on which harmonic has the highest level (1 = first harmonic, 2 = second harmonic, etc.).

Covered and open voices clearly differ in the strongest partial. Up to about 260–280 Hz (C4–C#4) the strongest partial is generally the third harmonic in covered voice and the fourth in open voice. Up to about 350 Hz (F#5) the second harmonic is the strongest in covered voice and the third in open voice. Above, it is generally the first harmonic in covered voice and the fourth in open voice. In the measures used in the present study, the maximum in covered voice is fairly systematically LowR up to C#4, then L2 between D4 and F4 and finally L1 above that. In contrast, the maximum in open voice is almost invariably the LowR resonance.

5.4.4 Synopsis of the results on covering

Systematic spectral differences were observed between the voice types investigated. As expected, piano voices differ from the others by their low dB values. Covered piano voice shows L1 dominance, whereas open piano voice resembles open forte voice by having higher L2, LowR and HighR values than covered piano voice. Forte voices differ much in voice range. When using open voice the singer was not able to produce the wide range that she was capable of in covered voice (and she felt open singing to be very strenuous and fatiguing). Open voice shows much amplitude in harmonics above the fundamental frequency. Covered voice, on the other hand, reveals the measured dB values in an intricate balance (shown in Figure 35 and interpreted above): LowR dominates in low notes, L2 in the area D4–F4 (294–349 Hz) and L1 above that. Figure 32 (on page 104) shows that L1 dominance continues beyond the area common to covered and open voice, i.e. above A#4 (466 Hz). The L1-L2 difference (see Figure 34) appears to reflect register transition: to keep the transition smooth, the difference is altered in the register transition area.

6 DISCUSSION

6.1 Spectral differences in the speaking and singing voice

6.1.1 Overview

The results of the analyses of the four sets of material relating to speaking and singing voice, presented above in summary form in Tables 5, 7 and 18 and in chapter 5.4.4, are compiled and contrasted in Table 22 (overleaf). The table shows the direction of statistically significant spectral differences in the subgroups in which the four sets of materials were analyzed. Three major spectral properties have been included: the difference between the levels of L1 and L2, the level at 1 and 2 kHz or at LowR and the level at 3 kHz and higher or HighR. In the supported and unsupported voice material, the results are given separately for low notes and high notes (Support/Low and Support/High in Table 22), as large differences were observed between them. Statistically non-significant but systematic differences in the dysphonia and vocal intensity materials have been included in parentheses.

The LTAS analysis revealed statistically significant differences between several groups of dysphonic speakers. In the cancer group the level of the second harmonic is high in relation to the first harmonic, while it is low in the paralysis group and in the slight dysphonia group. The male voices tend to have a higher level of L2 than the female voices, even though the difference in not statistically significant. The spectral level at 1–2 kHz is high in the severe dysphonia group, in the nodule group and in the hyperfunctional dysphonia group. The level is low in the paralysis group, in the slight dysphonia group and in the hypofunctional group. The level at high frequencies (3 kHz and above) is high in the cancer and nodule groups and in the severe dysphonia group, low in the paralysis group and in the slight dysphonia group and in the hypofunctional group. High levels at high frequencies may suggest the existence of turbulence in the voice. In sum, a shallow spectrum characterizes the cancer, nodule and severe dysphonia groups, a steep spectrum the paralysis and slight dysphonia groups. The observations are in line with previous studies (e.g. Hammarberg 1986,

TABLE 22 Significant differences in three spectral measures (L1-L2, 1 & 2 kHz or LowR and 3+ kHz or HighR) in all data sets. The groups indicated by ▲ have higher spectral amplitude, those by ▼ lower amplitude. In parentheses nonsignificant but systematic differences.

Data set	L1-L2	1 & 2 or LowR	3+ or HighR
▲			
Dysphonia	cancer (male)	severe, nodule (hyperfunctional) female	cancer, severe, nodule (hypofunctional) female
Vocal intensity	loud (male)	loud female	loud
Support/ Low	supported male		supported/male female
Support/ High	supported male	supported/male male	supported/male female
Covering	open forte	open forte	open forte
▼			
Dysphonia	paralysis, slight (female)	paralysis, slight (hypofunctional) male	paralysis, slight (hyperfunctional) male
Vocal intensity	soft (female)	soft male	soft
Support/ Low	unsupported female		unsupported/male male
Support/ High	unsupported female	unsupported/male female	unsupported/male male
Covering	covered piano	covered piano	covered piano

Kitzing 1986, Milenkovich, Bless & Rammage 1991), even though direct comparison is difficult due to differences in material and methods.

Clear and systematic differences between the three loudness conditions (loud, normal and soft) were revealed. The observations corroborate the results of previous studies (Fant 1959, Brandt et al. 1969, Gramming & Sundberg 1987, Gauffin & Sundberg 1989, Södersten et al. 1991). The spectral slope of loud voice is shallow and that of soft voice steep. The difference in slope is evident in many of the measures employed in the present study, including the difference in level between the second and first harmonic, the 1–2 kHz area and the area above 3 kHz.

Supported and unsupported voice differ in spectral properties. The difference in level of the second and first harmonic distinguishes supported and unsupported voice: both in low and high notes L2 is higher in supported voice and lower in unsupported voice. The other statistically significant spectral differences between supported and unsupported voice show up in male singers only: in low notes the level above 2 kHz (HighR) is higher in supported singing and lower in unsupported singing. In the high notes sung by men the levels at middle (.5–2 kHz) and high (2– kHz) frequencies are higher in supported singing and lower in unsupported singing. In the male singers, supported voices in general have a flatter spectrum and unsupported voices a steeper spectrum. The female singers as a group do not show significant differences at the middle and high frequencies. An analysis of each singer showed that the female spectra follow two patterns. One pattern resembles the male pattern: a shallow spectrum in supported singing, a steep spectrum in unsupported singing. The other pattern is opposite: a steep spectrum in supported voice and a shallow spectrum in unsupported voice. The existence of two patterns in the female singers accounts at least to some extent for the lack of significant differences between supported and unsupported voice in the female group. (The observed significant difference in L1-L2 is apparently due to those male singers and female singers who have a shallow slope in supported voice.)

Covered and open singing by the female singer investigated in this study differ in spectral properties. Covered voice has a lower overall amplitude than open voice. On the average (with all pitches pooled), covered voice shows the dominance of the fundamental and open voice dominance in the first formant area. These observations are in line with results by many researchers, for instance Pielke (1912) and Luchsinger (1951). They are also in line with comparisons between singers and nonsingers (e.g. Sundberg 1987); after all, when singing in open voice a singer largely simulates a nonsinger's singing. A more detailed analysis of each note sung in a pitch range common to both covered and open voice shows that in covered voice the levels of the low harmonics exist in intricate balance. In covered voice, the harmonics in the first formant area dominate in the low notes, the level of the second harmonic in the area D4–F4 (294–349 Hz), and the level of the fundamental above that. Open voices are dominated by the third, fourth and fifth harmonics, which in the samples are located in the region of the first and second formant. Thus, covered voice shows spectral dominance in lower harmonics than open voice. The dominance of the fundamental is evident also in piano covered voice, whereas piano open voice shows a much more level slope.

The spectral properties of the female and male voices differ both in the speaking voice and the singing voice materials. In dysphonic voices female subjects show higher dB levels both in the 1–2 kHz area and above 3 kHz. The level of L2 in relation to L1 is higher in male than female voices, even though the difference in level is not statistically significant. The spectral properties of voices produced at the prescribed sound pressure levels show differences between the female and male speakers. In relation to the highest peak, located below 1 kHz, the amplitude at 1–2 kHz is higher in the female than male speakers. On the other hand, the difference in level between the lowest two harmonics is not significant, but the level of L2 is higher in the male speakers than female speakers. Clear differences in this direction have been obtained by Södersten et al. (1991). The female–male differences are similar in dysphonic voices and in the study on voice at the prescribed sound pressure levels. The female speakers show a high amplitude at 1–2 kHz and male speakers have a higher level of L2 in relation to L1. The higher level at 1–2 kHz may be due to the lower formants being located higher in frequency in female than male voices. The more level spectral slope in male than female speakers indicated by the L1-L2 difference may be due to more adduction in male than female speakers.

In the study on support in singing, the female singers show a higher mean level than the male singers at the higher frequencies (above 2 kHz) both in low and high notes. In the middle frequencies (.5–2 kHz) the male voices show a higher mean level than the female voices, but only in the high notes. The mean level of the second harmonic in relation to the first harmonic is higher in the male voices than female voices both in the low and high notes. The L1-L2 difference follows the pattern observed in the results for the speaking voice. The differences observed between female and male singers in high notes in the middle frequencies may be due to the LowR measure being outside the low resonance (F1–F2) area in the female voices (as their pitch is rather high in high notes). The gender-related difference observed in the high frequencies is somewhat puzzling: as the singer's formant is typically a male phenomenon, higher levels of HighR would be expected in male than female singers. However, the results are opposite. An analysis of Figures 30 and 31 (on pages 90 and 91) shows that the higher level of HighR in female than male singers is mainly due to singer 5 in high notes and singers 5, 6 and 7 in low notes. The other female singers show values similar to male singers. Consequently, it is unclear whether the observed difference in mean HighR level between female and male singers can be interpreted as gender-specific. Besides, it is also possible that the higher mean level of HighR in female voices is due to turbulence in some, especially unsupported voices. Table 22 (on page 110) also shows that male singers singing in supported voice show a higher level in HighR than in unsupported singing. This is an expected result.

The directions of spectral differences presented in Table 22 are rather systematic. Table 23 divides the groups into those characterized by a high level of L1-L2, 1 & 2 or LowR and 3+ or HighR and those characterized by a low level in these measures.

The groups characterized by high spectral levels are the cancer and nodule groups as well as the hyperfunctional, loud, supported, open, forte and male voices. The groups characterized by low spectral levels are the paralysis group and the hypofunctional, soft, unsupported, covered, piano and female voices.

The low and high-level clusters, formed on the basis of the spectral measurements, appear to form meaningful wholes. The high-level cluster can tentatively be labelled "powerful", the low-level cluster "powerless". These clusters enter into the discussion of four central issues in the present study: normal and dysphonic voice, vocal intensity, singing voice, and female and male voice.

TABLE 23 Groups characterized by high and low levels of L1-L2, 1 & 2 or LowR and 3+ or HighR.

High levels	Low levels
cancer, nodule	paralysis
hyperfunctional	hypofunctional
loud	soft
supported	unsupported
open	covered
forte	piano
male	female

6.1.2 Normal and dysphonic voice

It is difficult to define a normal speaking voice as there is much variation in the voice of speakers who consider themselves normal and who are considered as normal. In terms of the spectral differences presented in Table 23, normal voices may be located somewhere in-between: they lack extreme spectral properties. Dysphonic voices on the other hand have such extreme properties. The LTA spectra of voices in the cancer group and in the paralysis group diverge in opposite directions from an imaginary baseline. However, it is true that there is much variation within both groups.

Cancer and paralysis of the recurrent nerve have direct and very different consequences on the vibration of the vocal folds: the cancer group is likely to show hyperadduction and the paralysis group hypoadduction. The groups represent two extremes, characterized by a shallow and steep spectral slope, respectively. Both have a connection with voice production: it is possible and even probable that the development of a growth in the larynx causes hyperadduction and a paralysis of the vocal folds causes hypoadduction. The results on dysphonic voices are as expected. They take a middle position between overly sceptic studies (e.g. Wendler, Rauhut & Krüger 1986) and hubristic studies claiming to distinguish minute shades of vocal quality by means of acoustic analyses.

The acoustic consequences of hyperadduction vs. hypoadduction are manifested in the ratio of the levels of the two lowest harmonics as well as in the level of the harmonics in the 1–2 kHz region (e.g. Lindqvist Gauffin 1972, Stevens 1977). Above that frequency, noise components may enter the spectra. Tentative evidence comes from the hypofunctional and hyperfunctional groups, which behave in a dual manner: in the 1–2 kHz region, the spectral amplitude is on the average higher in hyperfunctional voices than in hypofunctional voice, whereas at the higher frequencies (3 kHz and above) hypofunctional voices have more amplitude than hyperfunctional voices. The greater amplitude in hypofunctional voices at the higher frequencies may reflect turbulence produced in the relatively abducted glottis.

The practical utility of being able to acoustically distinguish voice groups formed on the basis of diagnosis and/or perceptual evaluation by long-term spectrum analysis may not be great. After all, a voice can often be evaluated rather accurately by listening and the cause of certain voice disorders can be determined by a clinical examination. However, acoustic methods such as LTAS analysis may be useful in monitoring the results of voice therapy. For instance, Kitzing and Åkerlund (1993) and Södersten and Hammarberg (1993) report an increase in amplitude levels, especially in the region of the first formant, after therapy. Such an increase is good for a weak, hypoadducted voice: the voice works better in communicative tasks where a loud voice is required. On the other hand, the effect of voice therapy on a hyperadducted voice should be the reverse in order to help detect finer shades of quality in a habitually strained loud voice.

The inclusion of spectral measures in phonetograms describing sound pressure level at selected fundamental frequencies (see e.g. Coleman et al. 1977) appears a useful endeavour. Attempts have been made to provide information on high-frequency noise in phonetograms. For instance, Pabon and Plomp (1988) included a measure for noise above 5 kHz. Gramming, Gauffin and Sundberg (1986) included the level of the fundamental in phonetograms. In view of the results obtained in the present study, the level of the fundamental or the difference in level between the fundamental and the second harmonic appear to be measures that adequately complement phonetographic measurements.

6.1.3 Vocal intensity

The present study and other studies have shown that voices produced at various sound pressure levels systematically differ in spectral slope. Changes in vocal intensity are accompanied by spectral modifications.

Why are human beings capable of such variation in spectral slope? Ostwald (1973) has given an answer:

> "We advance the hypothesis that speakers increase the amount of sound they emit at higher frequencies when they want to be listened to and reduce their high-frequency output when they prefer to be ignored."

Subsequent studies have demonstrated the appropriateness of this answer. For instance, Robinson and MacArthur (1982) have shown that a loud voice (75 dB, as compared to a voice 5 dB weaker) is more salient, gets more attention and is regarded as more causal. In general, high amplitude (and a shallow spectrum) is often associated with qualities like dominance, authority, credibility and extroversion (e.g. Scherer 1978), low amplitude (and a steep spectrum) with qualities like intimacy, insecurity and submission. Attributions such as dominance and submission apply not only in human beings but in (other) animals as well.

High-frequency output can be increased not only by phonating more loudly but also by shifting phonation toward more adducted, i.e. strained and creaky, phonation on the adduction–abduction continuum. Amplitude at high frequencies can be decreased by shifting phonation toward the more abducted, i.e. breathy, end of the continuum. Such shifts often accompany changes in vocal intensity, but they can also be controlled separately, as evidenced by the possibility of a loud breathy voice and a strained soft voice.

Listening tests have demonstrated that human beings are able to differentiate loudness and vocal effort (e.g. Brandt et al. 1969). One explanation for this ability has been given above: vocal effort is manifested in spectral properties, as a shallow or steep spectral slope.

The loud voices and the voices of the cancer group in the present study share spectral properties, as do the soft voices and the voices of the paralysis group. Does this mean simply that the cancer patients spoke in a louder voice and paralysis patients in a softer voice? It is a possibility, but the question cannot be examined more closely, as the sound pressure level was not registered while making the tape-recordings.

When making audio recordings for the examination of voice disorders it is important to register the sound pressure level of each patient, as abnormal or inappropriate loudness may be part of the disorder. This is a routine procedure in some clinics (Hammarberg 1985). Kitzing (1986) has proposed that the overall voice intensity should be kept under control in clinical applications of LTAS. This is advantageous for the analysis by standardizing the voice samples and facilitating comparison between them. On the other hand, the use of a single prescribed level may cause some patients to speak unnaturally, thereby degrading the data.

Speaking at prescribed (e.g. 55, 65, 75 dB) or subjectively chosen (e.g. soft, normal, loud) sound pressure levels may give more information about the voice than speaking at habitual pitch. Holmberg, Hillman, Perkell and Gress (1992:115) strongly advocate the use of multiple baselines, i.e. recording speech at several loudness levels, "in longitudinal single-subject studies of vocal function, to determine the extent to which observed parameter change is due to actual change of vocal behavior or to merely a possible effect of variation in SPL". Such variation in sound pressure level can yield information on the "vulnerability" of voice; speaking in a loud voice, for instance, may reveal voice problems that would not show up when speaking with habitual loudness. Voice fatigue may not show up for hours when speaking with a normal or soft degree of loudness.

6.1.4 Support in the singing voice

In a study on support in singing, based on the same recordings as the present study, Sonninen, Hurme & Sundberg (1994) reported preliminary observations on the difference between the levels of the first formant region and the first harmonic. According to the study, the difference in level is more positive in supported voice and more negative in unsupported voice; in other words, the spectral slope is shallower in supported voice and steeper in unsupported voice.

The results of the present study extend these observations by analyzing the levels of both the second harmonic and the first formant region. The results give a more complex picture: L2 is higher in relation to L1 in supported voice and lower in unsupported voice. The level of the first formant region is only higher in supported voice in the high notes sung by the male singers. Consequently, the spectral correlates of support in voice appear gender-related. Chapter 6.1.6 will further examine female–male differences in voice.

The results of the present study are also related to an extensive study on the physiological and acoustic characteristics of support in singing by Griffin et al. (1995). Their acoustic observations center on formant frequencies, which were not measured in the present study. However, they also report a higher amplitude of the fourth formant in supported singing than in unsupported singing and associate this observation with the singer's formant (see e.g. Sundberg 1974). In the present study, the male singers showed a difference in level at the higher frequencies (above 2 kHz): the level is high in supported singing and low in unsupported singing, which is in line with the results of Griffin et al. (1995). They also observed a high level of subglottal pressure (see also Gauffin & Sundberg 1989) and a higher sound pressure level in supported singing; louder voice has a more level spectrum, as shown above.

In the present study, perceptual ratings of the quality of supported and unsupported voices were related to the spectral measurements. In respect to all singers, a low level at low resonances (LowR, usually in the first and second formant region) correlates with good quality both in low and high notes. The supported samples produced by the female singers show a negative correlation: a lower level of L1-L2 is associated with good quality. These correlations establish tentative links between spectral properties and voice perception.

A closer analysis shows the female singers to cluster in two groups: One group consists of those who produce rather s small spectral difference between the low and high notes, even though the high notes receive somewhat higher ratings. The other group is characterized by rather low ratings in the low notes and much higher ratings in the high notes. The groups are partly the same but not identical to the two groups of female singers identified in the analysis of spectral differences: those with a "powerful" pattern similar to that of the male singers and those with a "powerless" pattern. Voice class does not explain the division of the female singers into two groups, as all are sopranos. However, it is possible that the subdivisions in female voice may reflect such *Fach*-related voice classifications as dramatic and lyric soprano. More research is needed to better understand such classifications.

6.1.5 Covering in the singing voice

In the data set of covered and open voice, the singer performs the transition from chest voice to a higher register in a very different manner in covered voice compared to open voice. The difference in level between the second and first harmonics (L1-L2) distinguishes covered and open forte voice in this singer. In covered voice, there is a turning point at about 350 Hz, closely corresponding to the subjective location of the transition at about E4–F4 (330–347 Hz). Above the turning point, L2 decreases in covered voice but continues to rise in relation to L1 in open voice. This way L1 dominates the spectrum above the transition in covered voice, but in open voice the higher harmonics dominate. In other words, covered voice has characteristics of modal voice up to the turning point; above that it has characteristics of the falsetto (see e.g. Sundberg 1977). On the other hand, open voice is modal (or even strained) throughout, until the relatively low pitch above which it cannot be produced. Covering is not only a way of crossing the transition area around E4, but also a way of continuing to raise pitch higher than the A#4 (466 Hz) area, without noticeable or abrupt changes in quality.

The differences observed in the spectra of covered and open voice may stem from two sources: laryngeal and supralaryngeal. Titze (1994a) describes two theories of equalization in register transition: register equalization with laryngeal adjustments and lung pressure and register equalization by vocal tract adjustments. They can both be applied to covered and open voice.

The first theory is physiological. Acoustic differences observed between covered and open voice can be accounted by laryngeal behavior: there is weaker contraction in the thyroarytenoid (vocalis) muscle in covered voice than in open voice (Sonninen et al. 1992). Differences have also been observed in the position of the larynx, due to the external muscles. Sonninen and Hurme (1996) report measurements of the laryngeal distances of the singer who sang the voice samples analyzed acoustically in chapter 5.4. Covered and open voice differ especially at E4–F#4 (330–370 Hz). In open voice, the larynx moves in a superior direction up to D#4 and from E4 to F#4 in an anterior-superior direction. The highest pitch in open voice was A#4, which the singer felt to be extremely forced and fatiguing. Covering made it possible to sing higher pitches. Covering was manifested biomechanically as blocking the superior-anterior movement (from D4 to E4) and starting again near the rest position (F4). At very high pitches (about 17 semitones higher than in open voice) the position of the larynx in covered voice was as superior and as anterior as in the highest (but much lower) pitches produced in open voice.

The study by Sonninen and Hurme (1996) demonstrates that covered and open singing differ in outer forces: open singing is characterized by excessive laryngeal tension and even vocal misuse, whereas covered singing is characterized by the skilful co-ordination of the muscles participating in the production of voice. A smooth register transition can be achieved by gradually decreasing the contraction in the vocal folds (Hirano 1981b, Sonninen et al. 1992). A consequence of reducing contraction in the vocal folds is less adduction. Thus, more abduction and dominance of the fundamental is to be expected when passing from chest voice to a higher register in covered voice.

The second theory is acoustic (see Titze 1994a). Basically, it accounts for register equalization by vocal tract adjustments, by tuning the formants in such a way that a smooth transition results (e.g. Miller & Schutte 1994). In singing, vowel quality can to some extent be "sacrificed" for voice quality. Formant frequencies can within certain limits be varied beyond what is acceptable in speaking in an attempt to make vocal timbre better. By shifting a formant to a harmonic or away from a harmonic, the level of the harmonic can be regulated. The adjustment of formants or formant tuning can increase or decrease the level of the harmonics and this possibility can be used in register equalization.

In the present study, the third harmonic generally has the highest amplitude from F#3 to C#4 (185–277 Hz) in covered voice; the first formant is apparently located in the area of the third harmonic (three times the fundamental frequency, i.e. 555–831 Hz). Between D4 and F4 (294–349 Hz) the second harmonic has the highest level; the first formant is apparently located in the area of the second harmonic (two times the fundamental frequency, i.e. 588–698 Hz). Above that area, fundamental frequency is too low (370–466 Hz) and the second harmonic too high (740–932 Hz) to "carry" the first formant without changing the vowel to something other than [ɑ]. This line of thinking gives an acoustic explanation for the turning point observed in covered voice. An acoustic explanation for the lack of a turning point in open voice would be that the formants are allowed to shift to the second (and higher) harmonics, thereby giving the vowel an [æ] quality.

Formant tuning may contribute to the differences observed in the present study between covered and open voice. To study formant tuning, an obvious method would be to measure formants. However, the measurement of formants in vowels with high fundamental frequency is notoriously difficult: there are too few harmonics to "carry" the formants. Another method of investigating formant tuning is perceptual. As a formant relatively close to the fundamental increases the level of the fundamental (Fant 1959, 1993), the vowel [ɑ] sung in the present study would tend to take on the color of [ɔ] in covered voice. Unpublished listening tests by Aatto Sonninen at the time when the recordings were made lend support to the idea that some formant tuning took place: several listeners reported impressions of the color of [ɔ] in some vowels produced in covered voice and impressions of the color of {œ} in some vowels in open voice.

The relative contribution of laryngeal and supralaryngeal factors has been estimated by Hertegård et al. (1990:226). According to them, less than one third of the observed difference in the level of the fundamental between covered and open voice can be explained by the somewhat lower first formant frequency in covered singing; the rest of the difference they interpret as due to changes in the voice source.

The present study cannot decide between the laryngeal and supralaryngeal theories of register equalization. It is clear that covered singing requires an intricate balancing of many factors, a skill mastered by professional singers. Singers learn to avoid "screaming", resulting from excessive tension in the vocal folds or from inadequate formant tuning.

6.1.6 Female and male voice

The dominance of the fundamental in the female voice has been observed in the present study and in many other studies of both dysphonic and non-dysphonic speakers (e.g. Hammarberg et al. 1984, Klatt & Klatt 1990). The prominence of the voice fundamental has also been observed in a number of studies of breathy voice (e.g. Ladefoged & Antoñanzas-Barroso 1985, Kasuya & Ando 1991). Indeed, the female voice is often characterized as breathy.

The existence of systematic differences between female and male voices does not prove that such differences are only due to anatomical differences in the vocal organs of women and men. Anatomical differences obviously contribute to voice differences, but the anatomical differences in female and male larynges and vocal tracts need not be great. After all, the members of both sexes vary much in size and also therefore in the size of their vocal organs. It is evident that the female voice and the male voice are partly a result of learning and socialization (Henton & Bladon 1985, Henton 1989, Günzburger 1991). Women want to or are expected to sound like women and men like men. A compelling argument for this view is the wide variation observed in different cultures in female and male voices in characteristics such as pitch and voice quality, both in speaking and in singing (Lomax 1978, Sonninen 1987, Ladefoged & Maddieson 1996).

Voice is breathy, if the vocal folds do not close completely during the closed phase of vocal fold vibration. An incomplete closure of the vocal folds is more common in females than males (Bless et al. 1986, Södersten 1994). The higher pitch of women also contributes to breathiness: the fundamental and the first formant tend to be closer to each other in the female voice than in the male voice, contributing to an increase in the level of the fundamental (Fant 1959).

In relation to singing voices, the acoustic results of supported and unsupported singing obtained in this study suggest that support in male singers may differ from that in female singers. This is in line with the results of Griffin et al. (1995). When contrasting perceptual ratings of the supported and unsupported voice samples with the acoustic measurements in chapter 5.3.2, it was found that male voices received better ratings if the level of the second harmonic was high in relation to the fundamental (L1-L2) and lower ratings if the level of L1-L2 was low. Female voices received better ratings if the L2 level was low in relation to L1, i.e. the fundamental was dominant, and lower ratings if the level of L1-L2 was high. According to the raters, female singers had a better voice quality if the fundamental was dominant, and male singers if the fundamental was weak in relation to the second harmonic.

Studies in person perception have shown that individuals are quickly and "automatically" assigned to categories such as female/male and young/old. Voice perception is always person perception: voices are never judged as such (except possibly in listening tests with drastically manipulated stimuli). When listening to a voice, it is categorized, for instance, as a female voice or a male voice, or as a young voice or an old voice. One category proposed in voice perception is that of the voice prototype. Voice prototypes act as templates when perceiving voices. (See Papçun, Kreiman & Davis 1989, Hosman 1989, Valo 1994.) For instance, a "feminine" female voice and a "masculine" male voice may act as prototypes against which perceived voices are matched (naturally in accordance with the ideals of each culture). The correlations between the quality ratings and the spectral properties of supported and unsupported voices reported in the present study indicate a gender-specific difference: female voices with a prominent voice fundamental are rated as good, male voices with a low level of the fundamental are rated as good. Are these among the characteristics of prototypical female and male voices?

In Table 23 (on page 113) supported and open singing are included in the "powerful" high-level cluster and unsupported and covered singing in the "powerless" low-level cluster. This appears inconsistent, as supported and covered singing would usually be characterized as good and unsupported and open singing as poor by singers and vocal pedagogues. The inconsistency may be resolved by a closer examination of the female voices. The spectral properties of covered voice resemble those of some of the female singers singing in supported voice; that is, the female singers not showing spectral characteristics similar to male singers. Thus, low spectral levels characterize covered voice (sung by a female singer), and supported voice as sung by the "not-male-like" female singers. High spectral levels characterize open voice (sung by a female singer) and supported voice as sung by male singers and "male-like" female singers. Consequently, the spectral characteristics of support cannot be generalized to singers in general. They seem to be different in female and male singers and they also differ across female singers.

The above discussion leads to an appraisal of the similarities in covering and support in the results of this study. A low spectral level in relation to the point of reference characterizes both the supported voice of "not-male-like" female singers and covered singing in the female singer investigated. Would covered singing by a male singer be characterized by a high or a low spectral level? In a study of covered and open voice in male singers, Hertegård et al. (1990) obtained results similar to those of the female singer investigated in the present study: in covered voice the level of the fundamental was higher and those of formants lower than in open voice. On the other hand, high spectral levels characterize both the supported voice of "male-like" female singers and male singers. Thus, covering and support may have more in common in female singers than in male singers. It is tempting to speculate that female singers have a "powerless" spectrum both in covered and supported singing, whereas male singers have a "powerful" spectrum in supported singing but a "powerless" spectrum in covered singing. Further research is needed to find out if this is true.

Such a discussion of the differences between female and male voice allows for an admittedly speculative synthesis of the results on the singing voice, arranged on the breathy–creaky continuum. Flow phonation can be placed rather close to normal phonation but leaning toward breathy phonation. Covered voice more or less equals flow phonation and open voice is close to normal phonation or perhaps leaning somewhat toward strained phonation. Supported singing by "not-male-like" female singers is similar to covered voice. Supported voice by male singers and "male-like" female singers may be closer to normal voice. The place of unsupported voice on the continuum is unclear, due to much variation in the unsupported singing. It may lie toward the strained end.

6.2 Evaluation of the study

The present study is broad in scope. Aspects of both the speaking voice and the singing voice have been studied and answers to the research questions obtained. In some cases the answers are tentative, as there is much variation in the voices investigated and as some of the groups are rather small (or consist of one subject, as in the covered and open voice samples). The research instruments used necessarily reflect the level of voice analysis technology at the time when the original analyses were made. The present study has accumulated over a relatively long period of time and was not originally planned as a whole. It may therefore lack in coherence.

The ecological validity of the present study is unsatisfactory: the speakers and singers did not produce voice in a natural setting. They were recorded in a clinical environment (the dysphonia and the covered/open samples) and in a laboratory environment (the vocal intensity and the supported/unsupported samples). This is unfortunate from the point of view of ecological validity, but understandable in an acoustic study of voice, as this kind of material would be difficult to record in a natural setting. For instance, speaking at prescribed sound pressure levels is only possible in the laboratory.

Measurements

The overall sound pressure level was registered during the production of covered and open singing. In the vocal intensity material (with prescribed sound pressure levels), the actual sound pressure levels presumably are rather close to the target. In the other two areas reported (dysphonia and support), the sound pressure level of the signal is not known. Therefore, measures relative to a reference level were taken. The spectral measurements were carried out relative to the highest amplitude, which is usually located in the first formant region. As seen above, the fundamental also often has the highest amplitude. This leads to several reference levels or baselines depending on the spectral slope, including those mentioned above as well as the level of the second harmonic. Despite the fact that there are several baselines, the average dB values computed reveal the slope in the spectrum. The overall sound pressure level would have given useful information on the different voices, but its registration was not possible in respect of the dysphonia, vocal intensity and support materials.

It is not feasible to compare the dB levels of the spectra in the four sets of materials, as the samples were recorded and analyzed in varying circumstances with varying instruments. Therefore, only the direction of the differences in relation to the highest amplitude has been studied when comparing the groups.

The relative spectral measures were useful in distinguishing between groups. This is true of the nine spectral measures at 1–10 kHz, of the LM and LH measures related to the LL point of reference and of the levels of the harmonics related to the level of the fundamental. The measures at 5–10 kHz distinguished between groups of dysphonic speakers, but not much between the samples of soft, normal and loud voice. The L1, L2, LowR and HighR measures also discriminated between groups in the analysis of singing voices. However, the amount of variation in the groups was often great, indicating that the groups were not very homogeneous in their spectra.

In the studies on singing voice the upper measurement points (LowR and HighR) were flexible in the high notes. This is a consequence of the high pitches produced by some singers; as there were only four harmonics below 3 kHz, e.g. at 700, 1400, 2100 and 2800 Hz, LowR was measured at a higher point than the area defined for it (= .5–2 kHz). The effect of this imprecision is minor. On the other hand, the use of four, albeit flexible, measurement points facilitated comparison.

These rather complex measures were ultimately reduced to three: L1-L2, middle frequencies (consisting of 1 & 2 kHz and LowR) and high frequencies (consisting of 3+ kHz and HighR). These measures appear to be useful when contrasting the results of the four sets of materials. They concisely sum up the essential properties of the spectra.

The difference in level between the first and second harmonics (L1-L2) with the fundamental as the point of reference appears a particularly useful measure, as it distinguishes the various groups better than the other measures. In many previous studies the difference between the level of the fundamental and that of the first formant region has been used to describe the properties of spectra. In the present study, however, this measure is not useful, as the absolute sound pressure level is not known and, consequently, the point of reference may be either the first formant region or the fundamental.

The measurements of the speaking and singing voices are limited by the instruments that were available at the time. The limitations are more severe in the analyses of the singing voice. The use of an 8-bit digitizer restricts both the dynamic range (to about 42 dB, Keller 1992 & 1994) and the frequency range (to about 6 kHz). On the other hand, both ranges appear satisfactory for the purposes of the present study. Below, in the context of discussing the measurement of the speaking voice, various views on the required frequency range in voice analysis are considered.

Speaking voice

Voices are individual and so are dysphonias. They have many causes, only imperfectly captured by diagnoses. Therefore, the acoustic study of dysphonia by comparing groups formed on the basis of diagnosis and/or perceptual evaluation may not reveal clear-cut differences. This is partly true in the results reported in this study. Nevertheless, some groups differ distinctly from others. For example, the cancer group and the vocal fold paralysis group differ maximally from each other.

It is conceivable that disordered voices show even more variation than normal voices. Therefore, dysphonic voices make a good test for the LTAS method: if no spectral differences can be detected in dysphonic voices, no differences will probably be detected in normal voices either. Spectral differences between dysphonic groups were detected in the present study and have been detected in other studies as well (e.g. Yanagihara 1967, Gauffin & Sundberg 1977, Hammarberg 1986, de Krom 1994).

The frequency range in long-term spectrum analysis is often limited to about 5 kHz. There are good grounds for this. If there is energy in the spectrum above 5 kHz, it is usually noise. Much of the amplitude in the 3–5 kHz area may also result from unvoiced components. It is common knowledge that the most relevant information from the phonetic point of view is found in the lowest frequencies of the spectrum (up to about 4 kHz). However, higher frequencies may contain information that marks the identity or state of a speaker or the quality of a singer (see e.g. Naranjo, Lara, Rodríguez & García 1994).

When the long-term average spectrum analyses reported in this study were started, one of the goals was to investigate possible spectral differences in dysphonic voices at high frequencies. For that purpose, LTA spectra were calculated up to 20 kHz (Hurme 1980b, Sonninen & Hurme 1982). Statistically significant differences were found, but sources of error were observed in the procedure. The analogical recording technology of the time and the clinical environment in which the recordings were made were not adequate for the purpose. Stevens (1985) lists some criticisms of the use of a mean sound level above 5 kHz: ”[– –] the directivity of sound radiation, cross modes in vocal tract resonance and noise level and sound reflections in the recording room may easily enter as sources of error.”

However, the idea of using very high frequencies in the analysis of voice has proved to be correct. Several studies, carried out with condenser microphones and digital audio tape-recorders, have demonstrated the potential of ”ultra-high” frequencies in discriminating between voices (Kasuya, Ogawa, Mashima & Ebihara 1986, Shoji, Regenbogen, Yu & Blaugrund 1991). Naranjo et al. (1994) showed that speakers with vocal nodules had a higher amplitude level at both 6–10 kHz and at 10–16 kHz as compared with a normal control group.

In the other speaking voice material the voices produced in the three loudness conditions differed distinctly in their spectral properties. This indicates that the long-term average spectrum method succeeds in describing variation in voice of this kind. This task of speaking in different loudness conditions has been suggested by Kitzing and Åkerlund (1993) as a test of the LTAS method.

The use of the three prescribed sound pressure levels the subjects were instructed to speak at may have disturbed some subjects, who experienced difficulty in speaking at the loud level (which was certainly louder than appropriate in an anechoic room). However, most subjects had no difficulty in speaking at the prescribed levels.

Singing voice

When recording the material for support and covering, the singers were asked to sing with support or without support and in covered voice or open voice. A way to operationalize support or covering is to start from the sensation the singer has of singing in supported and covered voice. The sensation is based on tradition passed on through training. It appears reasonable to trust the judgment of the singers themselves in this matter: after all, who would know better?

The present study has described the spectral properties of supported and covered voice in comparison with unsupported and open voice. The results of the LTAS analyses show that the singers in general were able to differentiate between supported and unsupported voice as well as between covered and open voice.

In the samples of supported and unsupported singing, there is too much variation in the pitch of the low and high notes. The singers should have been more carefully instructed to sing at about the same frequencies, below and above the register transition area. Many singers produced low notes at about 200 Hz and high notes at about 400 Hz, but one singer produced the low notes at about 400 Hz and all the female singers but one produced the high notes at about 600 Hz. Caution should be exercized in generalizing the results on support in singing – support may mean very different behavior to different singers. In fact, some singers feel that singing well has nothing to do with "support" and resist the use of the term.

The results on covered and open singing are very precise and they can be related to physiological (and to rudimentary perceptual) observations. However, they are based on one singer only. Therefore, caution is required in extending the results to covered singing in general.

6.3 Future directions

The human voice has been extensively studied by many methods for a long time. However, research topics often appear to be biased in several ways. Voice qualities used in more formal speech have been studied more than the qualities used in everyday, casual speech, with its social and regional variation. Male voices have been investigated more than female voices. Classical singing has been studied more than the many other types of singing voices, with their cultural and musical genre variation. The human voice needs to be studied in all its modifications.

The speaking voice shows many kinds of variation. Voices regarded as normal vary and dysphonic voices vary. Voice varies greatly across cultures and contexts. What is regarded as normal voice in one language may sound dysphonic in another. What is normal voice in one context may sound artificial or ridiculous in another.

Voice also varies across and within individuals. Young and old voices differ and so do female and male voices. There may be much variation in an individual's voice during the working day, for instance. Therefore, normal voice would merit increased research effort; by studying normal voice and variation in it a baseline could be established against which dysphonic voices, for instance, could be compared. The baseline for normal voice should not be determined on the basis of laboratory speech. The recording of material for the study of normal voice should take place in natural environments where people talk. Such material would better represent normal speaking voice.

There are numerous possibilities for further acoustic studies of normal speaking voice. The present study has shown that vocal intensity is reflected in the spectral properties obtained by long-term average spectrum analysis. This information may be useful in an area of vocal behavior which constitutes an occupational hazard for many individuals: vocal stress or fatigue. In rehabilitation, an individual's voice is often tested (in addition to investigating the everyday sound environment). One method is to stress a person's voice for hours at a stretch and observe the changes in the voice. Another method, which is less time-consuming, is to make the person speak at various vocal intensities. The intensities can be prescribed, as in the present study, or elicited by descriptions such as soft and loud. A person's voice can also be examined and interpreted on the breathy–creaky continuum. Information on the spectral characteristics of voice spoken at various vocal intensities can be used for that purpose.

The present study has shown some groups of dysphonic voices to differ acoustically in the LTA spectra. However, further research is needed to better understand the acoustic characteristics of various dysphonic voices. What are the diagnostic groups that show distinct acoustic correlates? Clearly, acoustic studies need to be complemented by physiological and perceptual studies in order to arrive at a more comprehensive picture of dysphonia.

The acoustics of the singing voice has been examined in the present study, as indeed in many other studies. Such studies, however, have tended to concentrate on classical singing. However, there are many varieties of singing: many musical cultures, many voice classes such as soprano and bass, many characteristics related to job specialty such as coloratura and counter tenor as well as belting and grunge. An acoustic approach to their study would yield useful quantitative data on the singing voice. Naturally, physiological and perceptual studies are also needed.

The present study has revealed gender-specific differences in the speaking and the singing voice. The topic is interesting: To what extent are the differences due to anatomy and to what extent to social factors? To what extent are female–male differences in voice universal and to what extent culture-specific? Cross-cultural voice research can give answers to such questions.

In the acoustic study of voice variation, long-term average spectrum analysis is a possible method. It is not *the* method. It can give answers to questions connected with the radiated voice signal and to some extent, as shown in this study, on questions to do with the voice source. There are methods specifically designed for the study of the voice source, such as inverse filtering. It is often advantageous to combine physiological and perceptual methods with acoustic methods. As emphasized above, by combining various methods in the analysis of a topic, new insights can be gained.

The need for more specificity in the development of voice terminology is a recurrent theme in studies on voice. Impressionistic terms often lack a commonly accepted definition. The development of scientific terms presupposes research: voice terms can be defined unambiguously only if the phenomenon to be named has been described and understood from the perspectives of both production, perception and acoustics. Voice research and terminological work need to interact.

Theory formation in the field of voice research has its foundation in the distant past; it has been vigorous during the last few decades. At the present, there is still much that is not understood – not even such a "simple" phenomenon as vocal fold vibration. Undoubtedly, a more comprehensive theory of voice, better able to describe and explain the production, acoustics and perception of voice, will be developed in the future. The human voice in its potential for communication and expression remains a challenge for research.

YHTEENVETO

Ihmisääni on ilmaisuvoimainen: ääntä voidaan muunnella monin tavoin. Siten äänessä on paljon vaihtelua sekä henkilöiden välillä että samallakin henkilöllä mm. eri tilanteissa, tunnetiloissa ja vaikkapa terveenä ja sairaana. Äänessä on monia erilaisia ominaisuuksia, joista kolme on äänentutkimuksen kannalta tärkeää: korkeus, voimakkuus ja laatu. Tämä tutkimus kohdistuu pääasiassa äänen laatuun.

Tutkimuksessa pohditaan aluksi äänen perustaa ja olemusta sekä äänen tehtäviä. Lisäksi selvitetään käytännöllisiä ja tieteellisiä lähestymistapoja ääneen. Ääntä tarkastellaan kolmesta perspektiivistä: foneettisesta, logopedis-foniatrisesta ja laulupedagogisesta. Työn taustaksi esitetään synteesi äänentutkimuksen historiasta sekä tutkimuksen kolmesta keskeisestä alueesta, äänen tuottamisen, akustiikan ja havaitsemisen tutkimuksesta. Työn yhtenä lähtökohtana on äänihuulien toimintaa ja äänen tuottamista kuvaava adduktio–abduktio -jatkumo, jossa ääni vaihtelee narinasta (engl. creak) puristeisen kautta normaaliin ja edelleen vuotoiseen (breathy). Jatkumo suhteutetaan ääniakustiikkaan, joka on työn keskeinen näkökulma.

Tutkimustyön pohjaksi analysoidaan äänentutkimuksen menetelmiä ja erityisesti spektrianalyysin ja keskiarvospektrianalyysin periaatteita. Keskiarvospektrianalyysin mittaamis- ja tulkintamenettelyt ovat olleet vaihtelevia ja osin epäjohdonmukaisia. Tässä työssä punnitaan keskiarvospektrianalyysin etuja ja haittoja sekä eritellään mittaamis- ja tulkintamenettelyjä. Lisäksi arvioidaan aikaisempien tutkimusten tuloksia erilaisten äänten spektriominaisuuksista: tieto äänihäiriöiden, äänen voimakkuusasteiden, erilaisten laulutapojen sekä nais- ja miesäänen spektrikorrelaateista on ollut hajanaista ja puutteellista.

Tutkimuksen tavoitteena on selvittää ihmisäänen spektriominaisuuksia. Paljastaako keskiarvospektri äänten eroja? Millaisia mahdolliset erot ovat? Tarkemmat tutkimuskysymykset liittyvät työssä tutkittuihin aineistoihin: Miten äänihäiriöt näkyvät keskiarvospektreissä? Miten äänen voimakkuus heijastuu spektrissä? Miten laulutapa (tukilaulu, peittolaulu) näkyy spektrissä? Miten peittolaulun spektriominaisuudet ja kuuntelu-arvioinnin tulokset suhtautuvat toisiinsa? Miten nais- ja miesäänten spektrirakenne eroaa? Tutkimus pyrkii vastaamaan myös teoreettisempaan kysymykseen: mitä äänentuotosta voi päätellä spektrin perusteella? Lisäksi työssä käsitellään metodologisia kysymyksiä: Mitä etuja ja haittoja keskiarvospektrianalyysista on? Mitkä mittaluvut parhaiten kuvaavat äänen vaihtelua?

Tutkimuksessa tarkastellaan äänten ominaisuuksia keskiarvospektrianalyysin avulla. Keskiarvospektrit tuotettiin Brüel & Kjær 2031 -spektrianalysaattorin ja Signalyze-ohjelman avulla. Tutkimus jakautuu neljään osaan tutkimuskysymysten ja analysoitujen aineistojen perusteella.

Ensimmäinen osatutkimus kohdistuu häiriöiseen ääneen. Mukaan otettiin sekä toiminnallisia että rakenteellisia häiriöitä. Aineistoon (n = 87) valittiin kuusi häiriöryhmää joko kuunteluarvion (ryhmät 1–3) tai foniatrisen diagnoosin perusteella (ryhmät 4–6): 1) lievä, 2) kohtalainen, 3) vaikea äänihäiriö, 4) äänihuulikyhmyt ja -polyypit, 5) äänihuulihalvaus ja 6) kurkkusyöpä. Keskiarvospektrit mitattiin useilla tavoilla (kymmenen mittauspistettä 1 kHz:n välein, kolme kaistaa taajuuksilla 0–2, 2–5 ja 5–8 kHz sekä toisen ja ensimmäisen osasävelen voimakkuusero eli L1-L2). Yleistulos on, että vaikka spektreissä on paljon vaihtelua, keskiarvospektrianalyysi erottelee jossain määrin äänihäiriöryhmiä. Selkeimmin erottuvat toisistaan kurkkusyöpä- ja äänihuulihalvausryhmät: halvausryhmässä spektrin kaltevuus on jyrkin, syöpäryhmässä vähäisin. Kaikilla mittaustavoilla saatiin tilastollisesti merkitseviä eroja joidenkin ryhmien välille. Toisen osasävelen taso suhteessa perussävelen tasoon erottelee ryhmiä samantapaisesti kuin laajemmat mitat: spektri laskee jyrkimmin äänihuulihalvausryhmässä ja lähes yhtä jyrkästi lievien häiriöiden ryhmässä sekä nousee hieman syöpäryhmässä. Nais- ja miesäänet eroavat toisistaan eniten 1 ja 3 kHz:n kohdalla: naisäänissä spektrin taso on korkeampi. Muitakin, joskaan ei tilastollisesti merkitseviä eroja havaittiin: Naisäänissä toisen osasävelen taso suhteessa perussävelen tasoon on matalampi kuin miesäänissä. Hyperfunktionaaliset äänet muistuttivat spektriltään syöpä-ääniä ja hypofunktionaaliset äänet halvausääniä. Kaiken kaikkiaan voidaan todeta, että keskiarvospektriä kuvaavat mittaluvut erottelevat äänihäiriöitä, etenkin elimellisiä.

Toisessa osatutkimuksessa selvitetään voimakkuudeltaan eritasoisten äänten spektriominaisuuksia. Aineisto koostuu luentanäytteistä, jotka 10 lukijaa tuotti 55, 65 ja 75 dB voimakkuudella, tarkkaillen samalla puheensa voimakkuustasoa desibelimittarista. Näytteet (n = 30) mitattiin keskiarvospektrianalyysin avulla. Tulokset ovat johdonmukaisia. Spektrit eroavat erityisesti 1–3 kHz:n alueella, jossa voimakkaan äänen spektri on normaaliin verrattuna korkeammalla tasolla ja hiljaisen äänen matalammalla. Naisäänissä on korkeampi spektritaso kuin miesäänissä 1 ja 5 kHz:n kohdalla. Vertailtaessa toisen osasävelen tasoa ensimmäisen osasävelen tasoon havaittiin, että normaaliin verrattuna voimakkaan äänen L2 on korkeammalla tasolla, hiljaisen äänen L2 matalammalla. Tutkittaessa viiden matalimman osasävelen voimakkuustasoja havaittiin, että nais- ja miesäänet eroavat systemaattisesti: naisäänissä spektrierot hiljaisen ja voimakkaan äänen välillä ovat suuremmat kuin miesäänessä.

Kolmas osatutkimus kohdistuu ns. tuettuun lauluääneen. Tuettu laulu on laulupedagogiikassa ja laulajien työssä tavallisimpia hyvän laulamisen määreitä. Tutkimuksessa halutaan selvittää, millaista tuettu laulu spektraalisesti on. Aineisto käsittää kahdeksan laulajan laulamia [pa]-tavujonoja, jotka on laulettu sekä matalalta että korkealta (rekisterirajan ala- ja yläpuolelta) sekä tuetusti että ilman tukea. Tukilaulu operationalisoitiin laulajien omaksi käsitykseksi ja toimintatavaksi. Ääninäytteitä tallennettaessa rekisteröitiin samalla useita fysiologisia muuttujia.

Laulunäytteistä laskettiin keskiarvospektrejä sekä tavujonojen alusta että lopusta. Niistä mitattiin ensimmäisen ja toisen osasävelen taso (L1, L2), "matalan" resonanssin taso (0.5–2 kHz, LowR) ja "korkean" resonanssin taso (2– kHz, HighR). Vaikka spektreissä on paljon vaihtelua, L1-L2-ero sekä matalissa että korkeissa vokaaleissa kertoo, että tuettuun ääneen liittyy pienempi spektrin kaltevuus, tukemattomaan ääneen suurempi. Toisin sanoen, tuetussa äänessä L1-L2-ero osoittaa loivempaa spektriä ja tukemattomassa äänessä jyrkempää spektriä. Lisäksi miesäänissä on tuetussa laulussa korkeampi spektritaso sekä LowR-alueella (korkeissa äänissä) että HighR-alueella (matalissa äänissä). Nais- ja miesäänien ero on selkeä kun tuettua ja tukematonta laulua ei erotella: L1-L2-ero osoittaa loivempaa spektriä miesäänissä kuin naisäänissä. Lisäksi miesäänissä on korkea spektritaso LowR-alueella. Naisääniä, jotka kaikki olivat sopraanoita, näyttää olevan kahta tyyppiä: toinen samantapainen kuin miesäänet (spektri siis suhteellisen loiva), toinen erilainen, ehkä naisäänelle ominainen (spektri jyrkkä).

Tuetut ja tukemattomat ääninäytteet arvioitiin myös kuuntelukokeessa. Arvioijia (laulajia, laulupedagogeja ja äänentutkijoita) oli 63. Käytössä oli kaksi arviointiasteikkoa, jotka ulottuivat nollasta sataan prosenttiin ja joihin merkittiin arvio tuen määrästä ja äänen laadusta. Koska tuki- ja hyvyysarvioilla oli vahva positiivinen korrelaatio, tulokset raportoidaan vain laadun hyvyys–huonous -skaalan osalta. Korkeat äänet arvioitiin keskimäärin paremmiksi kuin matalat. Tuetut äänet arvioitiin paremmiksi kuin tukemattomat. Naisäänet arvioitiin paremmiksi kuin miesäänet. Laatuarviot eivät eronneet sen mukaan, olivatko näytteet lauletun vokaalin alusta vai lopusta. Arvioinnit vaihtelivat suuresti lauluäänten välillä: korkeimmat arviot annettiin yhden naislaulajan korkealle tuetulle äänelle, matalimmat arviot toisen naislaulajan tuetulle matalalle äänelle. Naisäänet näyttävät laatuarvioinneissa jakautuvan kahteen ryhmään: ensimmäinen ryhmä sai melko hyviä arvioita sekä matalista että korkeista äänistä, toinen ryhmä puolestaan huonoja arvioita matalista äänistä ja hyviä arvioita korkeista äänistä. Ryhmät ovat vain osittain samat kuin akustisten mittausten perusteella muodostetut kaksi ryhmää, joten eroa ei voi selittää pelkästään spektriominaisuuksilla. Tuetun ja tukemattoman äänen arviointeihin vaikuttanevatkin myös mies- ja naislauluääneen kohdistuvat odotukset.

Suhteutettaessa tuettujen ja tukemattomien äänten kuunteluarvioinnit niiden spektrimittauksiin havaittiin korrelaatioita. Matalien äänten, miesäänten ja tukemattomien miesäänten saamat arvioinnit korreloivat L1-L2-eroon. Näiden äänten laatuarvio on parempi, jos toinen osasävel on korkealla voimakkuustasolla suhteessa ensimmäiseen osasäveleeseen (eli spektri on loiva). Tai kääntäen, laatuarvio on huonompi, jos L2 on matala suhteessa L1:een (eli spektri on jyrkkä). Toisaalta havaittiin korrelaatio L1-L2-eron sekä naisäänten ja tuettujen naisäänten saamien arviointien välillä. Laatuarvio on heikompi, jos L2 on korkea suhteessa perussäveleeseen (tai kääntäen, laatuarvio on parempi, jos L2 on matala L1:een verrattuna). Mies- ja naisäänet arvioidaan siis eri tavalla.

Neljäs osatutkimus käsittelee peitto- ja avolaulun spektraalisia ominaisuuksia. Näytteet koostuvat sopraanolaulajan simuloimasta voimakkaasta ja hiljaisesta peitto- ja avolaulusta ([ɑ]-vokaalia laulaen), josta laskettiin keskiarvospektrit. Näistä mitattiin ensimmäisen ja toisen osasävelen taso (L1, L2), "matalan" resonanssin taso (0.5–2 kHz, LowR) ja "korkean" resonanssin taso (2– kHz, HighR). Peittoisesti lauletussa hiljaisessa äänessä perussävel vallitsee, kun taas avoimesti lauletussa hiljaisessa äänessä toisen osasävelen taso sekä LowR ja HighR ovat tasoltaan korkeampia. Voimakkaasti lauletuissa vokaaleissa äänialue vaihteli suuresti: peittolaulussa alue oli erittäin laaja, kun taas avoäänen alue oli suppeampi (F#3–A#4 eli 185–466 Hz). Avolaulussa perussävelen yläpuoliset osasävelet ovat voimakkaita, erityisesti neljäs ja viides osasävel. Peittoäänessä taas vallitsee tarkka järjestys: matalissa äänissä vallitsee LowR, alueella D4–F4 (294–349 Hz) vallitsee toinen osasävel ja sitä korkeammalla taajuusalueella vallitsee perussävel. L1-L2-ero tuntuu kuvastavan hyvin peitto- ja avolaulun eroa.

Keskiarvospektrianalyysin tulokset voidaan suhteuttaa äänen tuottamiseen ja erityisesti ääniraon avautumista kuvaavaan adduktio–abduktio -jatkumoon. Tutkitut ääninäytteet eroavat yllä kuvatuilla tavoilla toisistaan ja muodostavat kaksi laajempaa joukkoa: jyrkän ja loivan kaltevuuden joukot. Jyrkän kaltevuuden joukkoon kuuluvat äänihuulihalvauspotilaiden ääni, hypofunktionaalinen ääni, hiljainen ääni, tukematon lauluääni, peittoinen lauluääni ja naisääni. Loivan kaltevuuden joukkoon puolestaan kurkkusyöpäpotilaiden ja äänihuulikyhmypotilaiden ääni, hyperfunktionaalinen ääni, voimakas ääni, tuettu ääni, avoin ääni ja miesääni. Jyrkän kaltevuuden ryhmä edustaa voimatonta, kantamatonta ääntä, loivan kaltevuuden ryhmä taas voimakasta, kantavaa ääntä. "Voimattomat" äänet sijoittuvat jatkumon abduktiosuuntaan, "voimakkaat" äänet adduktiosuuntaan. Erot johtunevat sekä anatomisista että sosiaalisista syistä. Esimerkiksi havaitut erot nais- ja miesäänessä selittyvät sekä mm. äänihuulten ja ääniväylän koosta että mm. oppimiseen ja sukupuolirooleihin liittyvistä tekijöistä.

Keskiarvospektri erottelee erilaisia ääniä. Kaikki käytetyt mitat (1–10 kHz, kolme taajuuskaistaa, LowR ja HighR sekä kahden alimman osasävelen – ja joissakin tapauksissa useampien osasävelten – voimakkuustasot) toimivat erottelutehtävässä. Erityisen systemaattisesti ääniä tuntuu erottelevan L1-L2-ero. Se vaikuttaa parhaimmalta adduktio–abduktio -jatkumoa kuvaavalta mitalta sekä alkuluvuissa esitetyn aikaisempien tutkimusten analyysin että nyt saatujen uusien mittaustulosten perusteella.

Tämä tutkimus kuvaa monenlaisia ääniä. Erot äänitysympäristöissä ja -laitteissa sekä absoluuttisen desibelitason puuttuminen kolmesta ensimmäisestä aineistosta aiheuttavat kuitenkin sen, ettei ääniä voi verrata toisiinsa muuten kuin kunkin aineiston sisällä. Näin ei esimerkiksi häiriöisiä ja normaaliääniä voida vertailla. Käytetyt tutkimuslaitteet kuvastavat akustisen äänentutkimuksen kehitystä viimeisen runsaan 10 vuoden aikana.

Äänen spektraalisten ominaisuuksien mittaaminen on yksi tapa kuvata äänen vaihtelua. Muita tapoja ovat akustisista menetelmistä esimerkiksi käänteissuodatus, perkeptuaalisista menetelmistä esimerkiksi äänen tutkiminen äänivaikutelmien kautta ja äänen tuottamisen tutkimusmenetelmistä esimerkiksi äänihuulten toiminnan seuraaminen elektroglottografian avulla. Rajoituksistaan huolimatta keskiarvospektrianalyysi on käyttökelpoinen menetelmä ihmisäänen kuvaamiseen.

Äänentutkimuksen yhtenä tavoitteena on ihmisääneen liittyvän variaation ja ääneen vaikuttavien tekijöiden selvittäminen. Äänentutkimuksen tuloksista on itsestäänselvää käytännön hyötyä. Tuloksia ja niiden sovelluksia odotetaan äänihäiriöistä kärsivien hoitoon ja terapiaan, puheammattilaisten ääniongelmien ehkäisyyn, laulupedagogiikkaan sekä vaikkapa teatteriin ja muuhun esiintymiseen liittyvään opetustyöhön. Kuitenkin äänentutkimuksen alalla tarvitaan kipeästi myös perustutkimusta ja sen pohjalle rakennettavaa äänen teoriaa. Erityisesti kaivataan äänen tuottamisen, ääniakustiikan ja äänellisten ominaisuuksien havaitsemisen välisten yhteyksien selvittämistä. Kokonaisuuden hahmottamisessa on äänentutkimuksen tulevaisuuden haaste.

BIBLIOGRAPHY

Abercrombie, D. 1967. Elements of general phonetics. Edinburgh: Edinburgh University Press.

Ågren, K. & J. Sundberg 1978. An acoustic comparison of alto and tenor voices. Journal of Research in Singing 1, 26-32.

Alku, P. & E. Vilkman 1995. Comparing methods for quantifying the voice source of different phonation types inverse-filtered from acoustic speech signals. In K. Elenius & P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, 13-19 August 1995, Vol. 2, 422-425.

Anders, L., H. Hollien, P. Hurme, A. Sonninen & J. Wendler 1988. Perception of hoarseness by several classes of listeners. Folia Phoniatrica 40, 91-100.

Arnold, G. 1973. Research potentials in voice registers. In J. Large (ed.) Vocal registers in singing, 137-153. The Hague: Mouton.

Aronson, A. 1980. Clinical voice disorders: An interdisciplinary approach. New York: Brian C. Decker.

Askenfelt, A. & B. Hammarberg 1986. Speech waveform perturbation analysis: A perceptual-acoustic comparison of seven measures. Journal of Speech and Hearing Research 29, 50-64.

Baken, R. 1987. Clinical measurement of speech and voice. London: Taylor & Francis.

Baltaxe, C. 1991. Vocal communication of affect and its perception in three to four year old children. Perceptual and Motor Skills 72, 1187-1202.

Barnes, J. (ed.) 1984. The complete works of Aristotle, Vol. 1. Princeton University Press.

Bernthal, J. & D. Beykelman 1977. The effect of changes in velopharyngeal orifice area on vowel intensity. Cleft Palate Journal 14, 63-77.

Bezooijen, R. van 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. Language and Speech 38, 253-265.

Bickley, C. 1982. Acoustic analysis and perception of breathy vowels. Working Papers (MIT Research Laboratory in Electronics, Speech Communications Group) 1, 71-82.

Bless, D., D. Biever & A. Shaik 1986. Comparisons of vibratory characteristics of young adult males and females. In S. Hibi, M. Hirano & D. Bless (eds.)

Proceedings of the international conference on voice, Vol. 2, 46-54. Kurume, Japan.

Blomberg, M. & K. Elenius 1970. Speech analysis. Statistical analysis of speech samples. Quarterly Progress and Status Report (Stockholm) 4, 1-8.

Bloothooft, G. 1985. Spectrum and timbre of sung vowels. Doctoral dissertation. Amsterdam: Vrije Universiteit.

Bloothooft, G. &. R. Plomp 1988. The timbre of sung vowels. Journal of the Acoustical Society of America 84, 847-860.

Brandt, J., K. Ruder & T. Shipp 1969. Vocal loudness and effort in continuous speech. Journal of the Acoustical Society of America 46, 1543-1548.

Bricker, P. 1965. Technique for objective measurement of speech levels. Journal of the Acoustical Society of America 38, 361-362.

Brown, B. & J. Bradshaw 1985. Towards a social psychology of voice variations. In H. Giles & R. St. Clair (eds.) Recent advances in language, communication, and social psychology, 144-181. London: Lawrence Erlbaum.

Bruyninckx, M., B. Harmegnies, J. Llisterri & D. Poch-Olivé 1994. Language-induced voice quality variability in bilinguals. Journal of Phonetics 22, 19-31.

Byrne, D., H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui, J. Kiessling, M. Kotby, N. Nasser, W. El Kholy, Y. Nakanishi, H. Oyer, R. Powell, D. Stephens, R. Meredith, T. Sirimanna, G. Tavartkiladze, G. Frolenkov, S. Westerman & C. Ludvigsen 1994. An international comparison of long-term average speech spectra. Journal of the Acoustical Society of America 96, 2108-2120.

Carlsson, G. & J. Sundberg 1992. Formant frequency tuning in singing. Journal of Voice 6, 256-260.

Carr, P. & D. Trill 1964. Long-term larynx-excitation spectra. Journal of the Acoustical Society of America 36, 2033-2040.

Castellengo, M., B. Roubeau & C. Valette 1985. Study of the acoustical phenomena characteristic of the transition between chest voice and falsetto. In A. Askenfelt, S. Felicetti, E. Jansson & J. Sundberg (eds.) SMAC 83, Proceedings of the Stockholm Acoustics Conference, July 28–August 1. 1983, Vol. 1, 113-123. Stockholm.

Catford, J. 1964. Phonation types. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott & J. Trim (eds.) In Honour of Daniel Jones, 26-37. London: Longmans.

Catford, J. 1977. Fundamental problems in phonetics. Edinburgh: University Press.

Coleman, R., J. Mabis & J. Hinson 1977. Fundamental frequency – sound pressure level profiles of adult male and female voices. Journal of Speech and Hearing Research 20, 197-204.

Colton, R. 1972. Spectral characteristics of the modal and falsetto registers. Folia phoniatrica 24, 337-344.

Colton, R. & H. Hollien 1972. Phonation range in the modal and falsetto registers. Journal of Speech and Hearing Research 15, 708-713.

Cooper, D. 1989. Voice: A historical perspective. Journal of Voice 3, 187-203.

Crandall, J. & O. MacKenzie 1922. Analysis of the energy distribution in speech. Physics Review 19, 222.

Damasio, A. & H. Damasio 1992. Brain and language. Scientific American, September 1992, 62-71.

Damsté, P. H. 1970, The phonetogram. Practica Oto-Rhino-Laryngologica 32, 185-187.

Daoud, S. 1965. Voice and speech in the writings of three medical writers of the 10th century: Rhazes, Haly-Abass, and Avicenna. In De therapia vocis et loquelae, Acta societatis internationalis logopediae et phoniatriae, XIII Congressus, Vol. II, 237-239. Wien.

Davis, S. 1979. Acoustic characteristics of normal and pathological voices. In Lass, N. (ed.) Speech and language, Volume 1, 271-335. New York: Academic Press.

Dejonckere, P. & G. Wieneke 1992. GRBAS-scaling of pathological voices: Reliability, clinical relevance and differentiated correlation with acoustic measures, especially with cepstral measurements. In E. Loebell (ed.) Proceedings of the XXIInd World Congress of the International Association of Logopedics and Phoniatrics, 3-15. Hannover: Medizinishe Hochschule.

De Krom, G. 1994. Acoustic correlates of breathiness and roughness: Experiments on voice quality. OTS Dissertation Series. Utrecht: Onderzoeksinstituut voor Taal en Spraak.

Dieroff, H. & C. Siegert 1966. Tonhöhenverschiebung unter Lärmbelastung. Folia Phoniatrica 18, 247-255.

Dmitriev, L. & A. Kiselev 1979. Relationship between the formant structure of different types of singing voices and the dimension of supraglottal cavities. Folia Phoniatrica 31, 238-241.

Dunn, H. & S. White 1940. Statistical measurements on conversational speech. Journal of the Acoustical Society of America 11, 278-288.

Elert, C.-C. & B. Hammarberg 1991. Regional voice quality variation in Swedish. Actes du XIIème Congrès International des Sciences Phonétiques, Vol. 4, 418-420. Aix-en-Provence: Université de Provence.

Esling, J. 1978. The identification of features of voice quality in social groups. Journal of the International Phonetic Association 8, 18-23.

Esling, J. 1982. Pronunciation considerations in ESL: Voice quality settings. Working Papers of the Linguistics Circle of the University of Victoria 2, 1-16.

Esling, J., B. Harmegnies & V. Delplancq 1991. Social distribution of long-term average spectra characteristics in Vancouver English. Actes du XIIème Congrès International des Sciences Phonétiques, Vol. 2, 182-185.

Estill, J., T. Baer, K. Honda & K. Harris 1985. Supralaryngeal activity in a study of six voice qualities. In A. Askenfelt, S. Felicetti, E. Jansson & J. Sundberg (eds.) SMAC 83, Proceedings of the Stockholm Acoustics Conference, July 28–August 1. 1983, Vol. 1, 313-322. Stockholm.

Fant, G. 1959. Acoustic analysis and synthesis of speech with application to Swedish. Ericsson Technics 1/1959, 1-108.

Fant, G. 1960. Acoustic theory of speech production. The Hague: Mouton.

Fant, G. 1973. Speech sounds and features. Cambridge, MA: MIT Press.

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

Fant, G. 1993. Some problems in voice source analysis. Speech Communication 13, 7-22.

Fant, G., J. Liljencrantz & Q. Lin 1985. A four-parameter model of glottal flow. Quarterly Progress and Status Report (Stockholm) 4, 1-13.

Fischer-Jørgensen, E. 1967. Statistic calculations of formant data. Annual Report (Institute of Phonetics, University of Copenhagen), 2, 158-170.

Fletcher, H. 1972. Speech and hearing in communication. Huntington, NY.

Fritzell, B. 1973. Foniatri för medicinare. Stockholm: Almqvist & Wiksell.

Fritzell, B., O. Hallén & J. Sundberg 1974. Evaluation of Teflon injection procedures for paralytic dysphonia. Folia Phoniatrica 26, 414-421.

Fry, D. 1979. The physics of speech. Cambridge: University Press.

Furui, S., F. Itakura & S. Saito 1972. Talker recognition by long-time average spectrum. Electronics and Communication in Japan 55, 54-61.

Gauffin, J. & J. Sundberg 1977. Clinical applications of acoustic voice analysis – acoustical analysis, results and discussion. Speech Transmission Laboratory. Quarterly Progress and Status Report (Stockholm) 2-3, 39-43.

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

Gauffin, J. & J. Sundberg 1989. Spectral correlates of glottal voice source waveform characteristics. Journal of Speech and Hearing Research 32, 556-565.

Glaze, L., D. Bless & R. Susser 1990. Acoustic analysis of vowel and loudness differences in children's voice. Journal of Voice 4:1, 37-44.

Gobl, C. & A. Ní Chasaide 1992. Acoustic characteristics of voice quality. Speech Communication 11, 481-490.

Gramming, P., J. Gauffin & J. Sundberg 1986. An attempt to improve the clinical usefulness of phonetograms. Journal of Phonetics 14, 421-427.

Gramming, P. & J. Sundberg 1988. Spectrum factors relevant to phonetogram measurement. Journal of the Acoustical Society of America 83, 2352-2360.

Gramming, P., J. Sundberg, S. Ternström, R. Leanderson, & Perkins, W. 1988. Relationship between changes in voice pitch and loudness. Journal of Voice 2, 118-126.

Griffin, B., P. Woo, R. Colton, J. Casper & D. Brewer 1995. Physiological characteristics of the supported singing voice. Journal of Voice 9, 45-56.

Günzburger, D. 1991. Breathiness in male and female speakers. Actes du XIIème Congrès International des Sciences Phonétiques, Vol. 3, 62-65.

Hammarberg, B. 1985. Clinical routines for the perceptual-acoustic assessment of dysphonia. Phoniatric and Logopedic Progress Report (Huddinge University Hospital, Sweden) 4, 14-29.

Hammarberg, B. 1986. Perceptual and acoustic analysis of dysphonia. Studies in Logopedics and Phoniatrics 1. Stockholm: Huddinge University Hospital.

Hammarberg, B. 1992. Pathological voice qualities – perceptual and acoustic characteristics of a set of Swedish "reference" voices. Phoniatric and Logopedic Progress Report (Huddinge University Hospital, Sweden) 8, 15-22.

Hammarberg, B., B. Fritzell, J. Gauffin, J. Sundberg 1986. Acoustic and perceptual analysis of vocal dysfunction. Journal of Phonetics 14, 533-547.

Hammarberg, B., B. Fritzell, J. Gauffin, J. Sundberg & L. Wedin 1980. Perceptual and acoustic correlates of abnormal voice qualities. Acta Otolaryngologica 90, 441-451.

Hammarberg, B., B. Fritzell & H. Schiratzki 1984. Teflon injection in 16 patients with paralytic dysphonia: Perceptual and acoustic evaluation. Journal of Speech and Hearing Disorders 49, 72-82.

Hammarberg, B. & J. Gauffin 1994. Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In O. Fujimura & M. Hirano (eds.), Vocal fold physiology: Voice quality control, 283-303. San Diego: Singular.

Harmegnies, B., J. Esling & V. Delplancq 1989. Quantitative study of the effects of settings changes on the LTAS. In J. Tubach & J. Mariani (eds.) Eurospeech 89. European Conference on Speech Communication and Technology, Vol. 2, 139-142. Paris.

Harmegnies, B., A. Landercy 1985. Language features in the long-term average spectrum. Revue de Phonétique Appliquée 73/74/75, 69-79.

Haskell, J. 1987. Vocal self-perception: The other side of the equation. Journal of Voice 1, 172-179.

Hattori, S., K. Yamamoto & O. Fujimura 1958. Nasalization of vowels in relation to nasals. Journal of the Acoustical Society of America 30, 267-274.

Henton, C. 1989. Fact and fiction in the description of female and male pitch. Language & Communication 9, 299-311.

Henton, C. & R. Bladon 1985. Breathiness in normal female speech: Inefficiency versus desirability. Language & Communication 5, 221-227.

Henton, C. & R. Bladon 1988. Creak as a sociophonetic marker. In L. Hyman & C. Li (eds.) Language, speech and mind. Studies in Honour of Victoria Fromkin, 3-29. London: Routledge.

Hertegård, S. 1994. Vocal fold vibrations as studied with flow inverse filtering. Studies in Logopedics and Phoniatrics 5. Stockholm: Huddinge University Hospital.

Hertegård, S., J. Gauffin & J. Sundberg 1990. Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. Journal of Voice 4, 220-230.

Hess, W. 1983. Pitch determination of speech signals: Algorithms and devices. Berlin: Springer.

Hirano, M. 1981a. Clinical examination of voice. Wien: Springer.

Hirano, M. 1981b. The function of the intrinsic laryngeal muscles in singing. In K. Stevens & M. Hirano (eds.) Vocal fold physiology, 155-170. Tokyo: University of Tokyo Press.

Hirano, M. 1988. Vocal mechanisms in singing: Laryngological and phoniatric aspects. Journal of Voice 2, 51-69.

Hirano, M. 1989. Objective evaluation of the human voice: Clinical aspects. Folia Phoniatrica 41, 89-144.

Hirano, M., Sh. Tanaka, M. Fujita & R. Terasawa 1991. Fundamental frequency and sound pressure level of phonation in pathological states. Journal of Voice 5, 120-127.

Hollien, H. 1984. Report on vocal registers. Acta Phoniatrica Latina 6, 11-22.

Holmberg, E., R. Hillman & J. Perkell 1988. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. Journal of the Acoustical Society of America 84, 511-529.

Holmberg, E., R. Hillman & J. Perkell 1995. Measures of the glottal airflow waveform, EGG, and acoustic spectral slope for female voice. I In K. Elenius

& P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, 13-19 August 1995, Vol. 3, 178-181.

Holmberg, E., R. Hillman, J. Perkell & C. Gress 1992. Relationships between SPL and aerodynamic and acoustic measures of voice production: Inter- and intra-speaker variation. Working Papers (MIT Research Laboratory in Electronics, Speech Communications Group) 8, 85-118.

Honikman, B. 1964. Articulatory settings. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott & J. Trim (eds.) In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961, 73-84. London: Longmans.

Hosman, L. 1989. The evaluative consequences of hedges, hesitations, and intensifiers. Human Communication Research 15, 383-406.

Huffman, M. 1985. Measures of phonation type in Hmong. Working Papers in Phonetics (UCLA) 61, 1-25.

Hurme, P. 1976. Imitation experiments in speech research: Data and theoretical considerations. In C.-Chr. Elert & T. Wictorsson (eds.) Symposium for Scandinavian Students on the Doctoral Level, 50-53. Umeå: Umeå University.

Hurme, P. 1980a. Auto-monitored speech level and average speech spectrum. In P. Hurme (ed.) Voice, speech and language: Reports and reviews, 119-127. Papers in Speech Research (University of Jyväskylä) 2. Publications from the Institute of Finnish Language and Communication 19.

Hurme, P. 1980b. Digitaalisen spektrianalyysin sovelluksia puheentutkimuksessa. In K. Suomi (ed.) Fonetiikan päivät – Oulu 1979, 40-49. Oulun yliopiston fonetiikan laitoksen julkaisuja 1.

Hurme, P. 1985. Äänen laadun määrittelemisestä ja kuvaamisesta. In O. Aaltonen & T. Hulkko (eds.) Fonetiikan päivät – Turku 1985, 65-84. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 26.

Hurme, P. 1986. Äänen laadun arvioimisesta ja mittaamisesta. In M. Lehtihalmes & A. Klippi (eds.) Logopedis-foniatrinen tutkimus Suomessa, 201-209. Suomen logopedis-foniatrisen yhdistyksen julkaisuja 19. Helsinki.

Hurme, P. & M. Pirinen 1984. Puheen keskiarvospektrien vaihtelusta. In U. Ikonen & T. Tikka (eds.) Fonetiikan päivät - Joensuu 1984, 23-36. Joensuun yliopisto. Kielitieteellisiä tutkimuksia 1.

Hurme, P. & A. Sonninen 1985a. Normal and disordered voice quality: Listening tests and long-term spectrum analyses. In P. Hurme (ed.) Papers in Speech Research (University of Jyväskylä) 6, 49-72. Publications of the Department of Communication 1.

Hurme, P. & A. Sonninen 1985b. Acoustic correlates of pathological voices: Individuals with paralysis of recurrent nerve. In P. Hurme (ed.) Papers in Speech Research (University of Jyväskylä) 6, 73-78. Publications of the Department of Communication 1.

Hurme, P. & A. Sonninen 1986. Acoustic, perceptual and clinical studies of normal and dysphonic voice. Journal of Phonetics 14, 489-492.

Hurme, P. & A. Sonninen 1995. Vertical and sagittal position of larynx in singing. In K. Elenius & P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, 13-19 August 1995, Vol. 1, 214-217.

Isshiki, N., H. Okamura, M. Tanabe & M. Morimoto 1969. Differential diagnosis of hoarseness. Folia Phoniatrica 21, 9-19.

Isshiki, N. N. Yanagihara & M. Morimoto 1966. Approach to the objective diagnosis of hoarseness. Folia Phoniatrica 18, 393-400.

Järvelä, E. 1991. Conveyed intention. A study of some acoustic aspects as related to production and perception of certain sung vowels. Doctoral dissertation. Helsinki: Sibelius Academy.

Javkin, H. & I. Maddieson 1983. An inverse filtering study of Burmese creaky voice. Working Papers in Phonetics (UCLA) 57, 115-125.

Johansson, C., J. Sundberg & H. Wilbrand 1985. X-ray study of articulation and formant frequencies in two female singers. In A. Askenfelt, S. Felicetti, E. Jansson & J. Sundberg (eds.) SMAC 83. Proceedings of the Stockholm International Music Acoustics Conference, Vol. 1, pp. 203-218. Publications Issued by the Royal Swedish Academy of Music, 46.

Karlsson, I. 1976. Vokalers spektrumlutning i tal. In C.-Chr. Elert & T. Wictorsson (eds.) Symposium for Scandinavian Students on the Doctoral Level, 68-72. Umeå: Umeå University.

Karlsson, I. 1992. Analysis and synthesis of different voices with emphasis on female speech. Doctoral dissertation. Royal Institute of Technology, Department of Speech Communication and Music Acoustics. Stockholm.

Kasuya, H. & Y. Ando 1991. Acoustic analysis, synthesis, and perception of breathy voice. In J. Gauffin & B. Hammarberg (eds.) Vocal fold physiology. Acoustic, perceptual, and physiological aspects of voice mechanisms, 251-258. London: Whurr.

Kasuya, H., K. Masubuchi, S. Ebihara & H. Yoshida 1986. Preliminary experiments on voice screening. Journal of Phonetics 14, 463-468.

Kasuya, H., S. Ogawa & Y. Kikuchi 1986. An acoustic analysis of pathological voice and its application to the evaluation of laryngeal pathology. Speech Communication 5, 171-181.

Kasuya, H., S. Ogawa, S. Mashima & S. Ebihara 1986. Normalized noise energy as an acoustic measure to evaluate pathologic voice. Journal of the Acoustical Society of America 80, 1329-1334.

Keidar, A., R. Hurtig & I. Titze 1987. The perceptual nature of vocal register change. Journal of Voice 1, 223-233.

Keller, E. 1992, 1994. Signalyze™, Signal Analysis for Speech and Sound, User's Manual. Program levels 2 and 3. InfoSignal™, C.P. 73, CH-1015 Lausanne, Switzerland.

Kelz, H. 1971. Articulatory basis and second language teaching. Phonetica 24, 193-211.

Kent, R. 1973. The imitation of synthetic vowels and some implications for speech memory. Phonetica 28, 1-25.

Kent, R. & C. Read 1992. The acoustic analysis of speech. San Diego, CA: Singular.

Kitzing, P. 1986. LTAS criteria pertinent to the measurement of voice quality. Journal of Phonetics 14, 477-472.

Kitzing, P. & L. Åkerlund 1993. Long-time average spectrograms of dysphonic voices before and after therapy. Folia Phoniatrica 45, 53-61.

Klatt, D. 1987. Acoustic correlates of breathiness: First harmonic amplitude, turbulence noise, and tracheal coupling. Journal of the Acoustical Society of America 82, 91.

Klatt, D. & L. Klatt 1990. Analysis, synthesis and perception of voice quality variations among female and male talkers. Journal of the Acoustical Society of America 87, 820-857.

Klingholtz, F. 1990. Acoustic representation of speaking-voice quality. Journal of Voice 4, 213-219.

Korhonen, M. 1993. Kielen synty. Helsinki: WSOY.

Kosiel, U. 1973. Statistical analysis of speaker-dependent differences in the long-term average spectrum of Polish speech. In W. Jassem (ed.) Speech analysis and synthesis, Vol. 3, 181-208. Warsaw: Polish Scientific Publishers.

Kreiman, J., B. Gerratt & K. Precoda 1990. Listener experience and perception of voice quality. Journal of Speech and Hearing Research 33, 103-115.

Kyttä, J. & P. Hurme 1982. Acoustical aspects of nasality. In P. Hurme (ed.) Vox humana, 203-211. Papers in Speech Research (University of Jyväskylä) 5. Publications from the Institute of Finnish Language and Communication 30.

Ladefoged, P. 1988. Discussion of phonetics: A note on some terms for phonation types. In O. Fujimura (ed.) Vocal physiology: Voice production, 373-375. New York: Raven.

Ladefoged, P. & N. Antoñanzas-Barroso 1985. Computer measures of breathy voice quality. Working Papers in Phonetics (UCLA) 61, 79-86.

Ladefoged, P., I. Maddieson 1996. The sounds of the world's languages. Oxford: Blackwell.

Ladefoged, P., I. Maddieson & M. Jackson 1988. Investigating phonation types in different languages. In O. Fujimura (ed.) Vocal physiology: Voice production, 297-317. New York: Raven.

Large, J. 1972. Towards an integrated physiologic-acoustic theory of vocal registers. The NATS Bulletin (February 1972), 18-36.

Large, J. 1973. Introduction [to vocal registers in singing]. In J. Large (ed.) Vocal registers in singing, 7-22. The Hague: Mouton.

Lass, N., D. Ruscello, L. Stout & F. Hoffmann 1991. Peer perceptions of normal and voice-disordered children. Folia Phoniatrica 43, 29-35.

Laukkanen, A.-M., E. Vilkman, P. Alku & H. Oksanen 1996. Physical variations related to stress and emotional state: A preliminary study. Journal of Phonetics 24, 313-335.

Laver, J. 1978. The concept of articulatory settings: An historical survey. Historiographia Linguistica 5, 1-14.

Laver, J. 1980. The phonetic description of voice quality. Cambridge: Cambridge University Press.

Laver, J. 1981. The analysis of voice quality: From the classical period to the twentieth century. In R. Asher & E. Henderson (eds.) Towards a history of phonetics, 79-99. Edinburgh: University Press.

Laver, J. 1994. Principles of phonetics. Cambridge: Cambridge University Press.

Laver, J., S. Hiller, J. MacKenzie & E. Rooney 1986. An acoustic screening system for the detection of laryngeal pathology. Journal of Phonetics 14, 517-524.

Laver, J. & P. Trudgill 1979. Phonetic and linguistic markers in speech. In K. Scherer & H. Giles (eds.) Social markers in speech, 1-32. Cambridge: Cambridge University Press.

Laver, J., S. Wirz, J. MacKenzie & S. Hiller 1981. A perceptual protocol for the analysis of vocal profiles. Work in Progress (Edinburgh) 14, 139-155.

Lee, Ch.-K. & D. Childers 1991. Some acoustical, perceptual, and physiological aspects of voice quality. In J. Gauffin & B. Hammarberg (eds.) Vocal fold physiology. Acoustic, perceptual, and physiological aspects of voice mechanisms, 233-242. London: Whurr.

Lehtonen, J. & P. Hurme 1980. The speech chain and the science of speech. In P. Hurme (ed.) Ääni. puhe ja kieli: selosteita ja katsauksia, 1-28. Puheentutkimuksen alalta 2. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 19.

Leino, T. 1994. Long-term average spectrum study on speaking voice quality in male actors. In A. Friberg, J. Iwarsson, E. Jansson & J. Sundberg (eds.) Proceedings of the Stockholm Music Acoustics Conference, July 28 – August 1, 1993, 206-210. Stockholm.

Leistiö, T. 1995. Alussa oli laulu. Tiede 2000 nro 6, 20-25.

Levman, B. G. 1992. The genesis of music and language. Ethnomusicology 36, 147-170.

Li, K.-P., G. Hughes & A. House 1969. Correlation characteristics and dimensionality of speech spectra. Journal of the Acoustical Society of America 46, 1019-1025.

Lieberman, P. 1963. Some acoustic measures of fundamental periodicity of normal and pathologic larynges. Journal of the Acoustical Society of America 35, 344-353.

Lindqvist Gauffin, J. 1972. A descriptive model of laryngeal articulation in speech. Quarterly Progress and Status Report (Stockholm) 2-3, 1-9.

Löfqvist, A. 1986. The long-time-average spectrum as a tool in voice research. Journal of Phonetics 14, 471-475.

Löfqvist, A. & B. Mandersson 1987. Long-time average spectrum of speech and voice analysis. Folia Phoniatrica 39, 221-229.

Löqvist, A. & L. Schalén 1991. Perceptual and acoustic analysis of the voice in acute laryngitis. Actes du XIIème Congrès International des Sciences Phonétiques, Vol. 4, 342-345. Aix-en-Provence: Université de Provence.

Lomax, A. 1978. Folk song style and culture. New Brunswick, NJ.

Luchsinger, R. 1951. Schalldruch- und Geschwindigkeitsregistrierung der Atemluft beim Singen. Folia Phoniatrica 3, 25.

Luchsinger, R. & G. Arnold 1965. Voice–Speech–Language. Clinical communicology: Its physiology and pathology. Belmont, CA: Wadsworth.

MacRecorder™ Sound System for the Macintosh™, 1991. Berkeley, CA: Farallon.

Majewski, W., H. Rothman & H. Hollien 1977. Acoustic comparisons of American English and Polish. Journal of Phonetics 5, 247-251.

Michel, J. 1985. The elusive nature of voice "quality". In Lawrence, V. L. (ed.) Transcripts of the 14th Symposium: Care of the Professional Voice, 178-183. New York: The Voice Foundation.

Milenkovic, P., D. Bless & L. Rammage 1991. Acoustic and perceptual characterization of vocal nodules. In J. Gauffin & B. Hammarberg (eds.)

Vocal fold physiology. Acoustic, perceptual and physiological aspects of voice mechanisms, 265-272. San Diego: Singular.

Miller, D. & H. Schutte 1994. Toward a definition of male 'head' register, passaggio, and 'cover' in Western operatic singing. Folia Phoniatrica Logopedica 46, 157-170.

Miller, R. 1959. Nature of the vocal cord wave. Journal of the Acoustical Society of America 31, 667-677.

Minifie, F., J. Green, J. Smith, & D. Huang 1995. Inter-judge variability in perception of voice qualities. In K. Elenius & P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, 13-19 August 1995, Vol. 2, 758-761.

Monsen, R. & A. Engebretson 1977. Study of variations in the male and female glottal wave. Journal of the Acoustical Society of America 62, 981-993.

Mörner, M. 1963. Bidrag till internationell nomenklatur för tonlägen och röstbeskrivning. Royal Institute of Technology (Stockholm), Speech Transmission Laboratory: Technical report 1/1963.

Murry, T. & E. T. Doherty 1980. Selected acoustic characteristics of pathologic and normal speakers. Journal of Speech and Hearing Research 23, 361-369.

Naranjo, N., E. Lara, I. Rodriguez, & G. García 1994. High-frequency components of normal and dysphonic voices. Journal of Voice 8, 157-162.

Ní Chasaide, A. & C. Gobl 1995. Towards acoustic profiles of phonatory qualities. In K. Elenius & P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, vol. 4, 6-13.

Nittrouer, S., R. McGowan, P. Milenkovic & D. Beehler 1990. Acoustic measurements of men's and women's voices: A study of context effects and covariation. Journal of Speech and Hearing Research 33, 761-775.

Nolan, F. 1983. The phonetic bases of speaker recognition. Cambridge: Cambridge University Press.

Nolan, F. &. B. Kühnert 1995. The articulatory correlates of descriptive categories for supralaryngeal voice qualities. In K. Elenius & P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, vol. 2, 470-473.

Nordström, P.-E. 1977. Female and infant vocal tracts simulated from male area functions. Journal of Phonetics 5, 81-92.

Ohala, J. 1981. The nonlinguistic components of speech. In J. Darby (ed.) Speech evaluation in psychiatry, 39-49. New York: Grune and Stratton.

Ohala, J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. Phonetica 41, 1-16.

Ohlsson, A.-C., B. Järvholm & A. Löfqvist 1987. Vocal symptoms and vocal behavior in teachers. Working Papers in Logopedics and Phoniatrics (Lund) 4, 36-45.

Ostwald, P. 1973. The semiotics of human sound. The Hague: Mouton.

Pabon, J. & R. Plomp 1988. Automatic phonetogram recording supplemented with acoustical voice-quality parameters. Journal of Speech and Hearing Research 31, 710-722.

Pabst, F. & J. Sundberg 1992. Tracking multi-channel electroglottograph measurement of larynx height in singers. Quarterly Progress and Status Report (Royal Institute of Technology, Stockholm), 2-3/1992, 67-78.

Panconcelli-Calzia, G. 1961. 3000 Jahre Simmforschung. Marburg: N. G. Elwert.

Papçun, G., J. Kreiman & A. Davis 1989. Long-term memory for unfamiliar voices. Journal of the Acoustical Society of America 85, 913-925.

Perelló, J. 1987. Phoniatrics chronology. Folia Phoniatrica 39, 153-166.

Peterson, G. & H. Barney 1952. Control methods used in a study of vowels. Journal of the Acoustical Society of America 24, 175-184.

Pielke, W. 1912. Über "offen" und "gedeckt" gesungene Vokale. Passow Schäfer Beiträge 5, 215.

Pittam, J. 1987. Listeners' evaluation of voice quality in Australian English speakers. Language and Speech 30, 99-113.

Pittam, J. 1994. Voice in social interaction. An interdisciplinary approach. London: Sage.

Pittam, J. & C. Gallois 1986. Predicting impressions of speakers from voice quality: Acoustic and perceptual measures. Journal of Language and Social Psychology 5, 233-247.

Pittam, J. & K. Scherer 1993. Vocal expression and communication of emotion. In M. Lewis & J. Haviland (eds.) The handbook of emotion, 185-197. New York: Guilford.

Prytz, S. & B. Frøkjær-Jensen 1976. Long-term average spectra analyses of normal and pathological voices. Folia Phoniatrica 28, 280.

Reich, W. 1950. Physikalische und somatische Voraussetzungen der Stimmbildung. Ciba-Zeitschrift 11 (123), 4502-4508.

Reimers Niels, L. & E. Yairi 1987. Effects of speaking in noise on vocal fatigue and vocal recovery. Folia Phoniatrica 39, 104-112.

Richman, B. 1993. On the evolution of speech: Singing as the middle term. Current Anthropology 34, 721-722.

Robinson, J. & L. Z. McArthur 1982. Impact of salient vocal qualities on causal attribution for a speaker's behavior. Journal of Personality and Social Psychology 43, 236-247.

Rossing, T. 1990. The study of sound. Reading, MA: Addison-Wesley.

Rothenberg, M. 1973. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. Journal of the Acoustical Society of America 53, 1632-1645.

Rothenberg, M. 1992. A multichannel electroglottograph. Journal of Voice 6, 36-43.

Sataloff, R. 1991. Professional voice. The science and art of clinical care. New York: Raven.

Sallinen-Kuparinen, A. 1978. Hur talar en finne? In E. Gårding, G. Bruce & R. Bannert (eds.) Nordic Prosody, 255-261. Lund: Lund University, Department of Linguistics.

Scherer, K. 1978. Personality inference from voice quality: The loud voice of extroversion. European Journal of Social Psychology 8, 467-487.

Scherer, K. 1979. Personality markers in speech. In K. Scherer & H. Giles (eds.) Social markers in speech, 147-209. Cambridge: Cambridge University Press.

Scherer, K. 1982. Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman (eds.) Handbook of methods in nonverbal behavior research, 136-198. Cambridge: Cambridge University Press.

Scherer, K. 1986. Vocal affect expression: A review and a model for future research. Psychological Bulletin 99, 143-165.

Scherer, R. & I. Titze 1987. The abduction quotient related to vocal quality. Journal of Voice 1, 246-251.

Schulman, R. 1989. Articulatory dynamics of loud and normal speech. Journal of the Acoustical Society of America 85, 295-312.

Seidner, W., H. Schutte, J. Wendler & A. Rauhut 1985. Dependence of the high singing formant on pitch and vowel in different voice types. In A. Askenfelt, S. Felicetti, E. Jansson & J. Sundberg (eds.) Proceedings of the Stockholm Music Acoustics Conference 1983, Vol. 1, 261-268. Stockholm: Royal Swedish Academy of Music.

Seidner, W. & J. Wendler 1982. Die Sängerstimme. Berlin: Henschelverlag.

Shipp, T. & K. Izdebski 1975. Vocal frequency and vertical larynx positioning by singers and nonsingers. Journal of the Acoustical Society of America 58, 1104-1106.

Shoji, K., E. Regenbogen, J. Yu & S. Blaugrund 1991. High-frequency components of normal voice. Journal of Voice 5, 29-35.

Sies, L. 1987. Voice and voice disorders. Sprinfield, IL: Charles C. Thomas.

Sivian, L. 1930. Speech power measurements. Journal of the Acoustical Society of America 2, 1.

Södersten, M. 1994. Vocal fold closure during phonation. Studies in Logopedics and Phoniatrics 3. Stockholm: Huddinge University Hospital.

Södersten, M. & B. Hammarberg 1993. Effects of voice training in normal-speaking women: Videoscopic, perceptual, and acoustic characteristics. Scandinavian Journal of Logopedics and Phoniatrics 18, 33-42.

Södersten, M., P.-Å. Lindestad & B. Hammarberg 1991. Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers. In J. Gauffin & B. Hammarberg (eds.) Vocal fold physiology. Acoustic, perceptual and physiological aspects of voice mechanisms, 217-224. San Diego: Singular.

Sondhi, M. 1975. Measurement of the glottal waveform. Journal of the Acoustical Society of America 57, 228-232.

Sonninen A. 1956. The role of the external laryngeal muscles in length-adjustment of the vocal cords in singing. Acta Otolaryngologica, suppl. 118, 218-231.

Sonninen, A. 1968. The external frame function in the control of pitch in the human voice. Annals of the New York Academy of Sciences, Vol. 155, 68-90.

Sonninen, A. 1987. Tarvitaanko Suomessa äänentutkimusta. Laulupedagogi 1987, 15-27.

Sonninen, A. 1990. Ihmisäänen rekistereistä. Laulupedagogi 1989-1990, 4-21.

Sonninen, A. 1993. Tukilaulun ongelma. Laulupedagogi 1993, 22-34.

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

Sonninen, A. & P. Hurme 1982. Clinical and acoustic observations of normal and hoarse voices. In P. Hurme (ed.) Phonetic aspects of normal and pathological speech, 1-16. Papers in Speech Research (University of

Jyväskylä) 4. Publications from the Institute of Finnish Language and Communication 29.

Sonninen, A. & P. Hurme 1984. Äänen laadun kuuntelukoe ja keskiarvospektrianalyysi: käheä ja normaali. In U. Ikonen & T. Tikka (eds.) Fonetiikan päivät - Joensuu 1984, 76-92. Joensuun yliopisto. Kielitieteellisiä tutkimuksia 1.

Sonninen, A. & P. Hurme 1992. On the terminology of voice research. Journal of Voice 6, 188-193.

Sonninen, A. & P. Hurme 1996. Control of vocal fold strain in singing: Roentgenographic and acoustic observations. Manuscript submitted for publication.

Sonninen, A., P. Hurme & J. Sundberg 1994. Physiological and acoustic observations of support in singing. In A. Friberg, J. Iwarsson, E. Jansson & J. Sundberg (eds.) SMAC 93, Proceedings of the Stockholm Music Acoustics Conference July 28 – August 1, 1993, 254-258. Publications issued by the Royal Swedish Academy of Music 79. Stockholm.

Sonninen, A., P. Hurme & E. Vilkman 1992. Roentgenological observations on vocal fold length-changes with special reference to register transition and open/covered voice. Scandinavian Journal of Logopedics & Phoniatrics 17, 95-106.

Sonninen, A., P. Hurme & R. Toivonen 1988. Computer voice fields in the study of normal phonation. Acta Phoniatrica Latina 10, 345-356.

Stemple, J. 1984. Clinical voice pathology: Theory and managament. Columbus: Charles E. Merrill.

Stevens, K. 1977. Physics of laryngeal behavior and larynx models. Phonetica 34, 264-279.

Stevens, K. 1985. Discussion comment on long-time average spectrum analysis, summarized by session chairman (Sundberg 1986) in the Voice Acoustics and Dysphonia conference in Gotland.

Sundberg, J. 1973. The source spectrum in professional singing. Folia Phoniatrica 25, 71-90.

Sundberg, J. 1974. Articulatory interpretation of the "singing formant". Journal of the Acoustical Society of America 55, 838-844.

Sundberg, J. 1977. Studies of the soprano voice. Journal of Research in Singing 1, 25-35.

Sundberg, J. 1984. Panel discussion on voice: loudness and projection. In V. Lawrence (ed.) Transcripts of the 13th Symposium Care of the Professional Voice, 235-238. New York: The Voice Foundation.

Sundberg, J. 1986. Chairman's summary of long-term average spectrum analysis section. Journal of Phonetics 14, 493-494.

Sundberg, J. 1987. The science of the singing voice. DeKalb: Northern Illinois University Press.

Sundberg, J. 1994. Perceptual aspects of singing. Journal of Voice 8, 106-122.

Sundberg, J. 1995. Subglottal pressure behavior in singing and speech. In K. Elenius & P. Branderud (eds.) Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, 13-19 August 1995, Vol. 3, 98-104.

Sundberg, J. & J. Gauffin 1979. Waveform and spectrum of the glottal voice source. In B. Lindblom & S. Öhman (eds.) Frontiers of speech communication research: Festschrift for Gunnar Fant, 301-320. London: Academic Press.

Sundberg, J., P. Gramming & J. Lovetri 1993. Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing. Journal of Voice 7, 301-310.

Sundberg, J., I. Titze & R. Scherer 1993. Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. Journal of Voice 7, 15-29.

Sweet, H. 1877. A handbook of phonetics. Oxford: Clarendon Press.

Talvi, W. 1931. Huru röstorganen behärskas. Stockholm: Kulturförlaget.

Tarneaud, J. 1961. Traité pratique de phonologie et phoniatrie. Paris: Maloine.

Tarnóczy, T. & G. Fant 1964. Some remarks on the average spectrum. Speech Transmission Laboratory. Quarterly Progress and Status Report (Stockholm) 4, 13-14.

Titze, I. 1988. A framework for the study of vocal registers. Journal of Voice 2, 183-194.

Titze, I. 1994a. Principles of voice production. Englewood Cliffs, NJ: Prentice Hall.

Titze, I. 1994b. Definitions and nomenclature related to voice quality. In O. Fujimura & M. Hirano (eds.) Vocal fold physiology: Voice quality control, 335-342. San Diego: Singular.

Titze, I. 1995. Speaking vowels versus singing vowels. Journal of Singing, September/October 1995. (An electronic version accessed at http://www.shc.uiowa.edu/research/nats/)

Titze, I. & J. Sundberg 1992. Vocal intensity in speakers and singers. Journal of the Acoustical Society of America 91, 2936-2946.

Trittin, P. & A. de Santos y Lleó 1995. Voice quality analysis of male and female Spanish speakers. Speech Communication 16, 359-368.

Trudgill, P. 1974. The social differentiation of English in Norwich. Cambridge: Cambridge University Press.

Valo, M. 1994. Käsitykset ja vaikutelmat äänestä. Studia Philologica Jyväskyläensia 33. Jyväskylän yliopisto.

Vilkman, E., P. Alku & A.-M. Laukkanen 1995. Vocal-fold collision mass as a differentiator between registers in the low-pitch range. Journal of Voice 9, 66-73.

Vilkman, E., A. Sonninen, P. Hurme & P. Körkkö 1996. External laryngeal frame function in voice production revisited: A review. Journal of Voice 10, 78-92.

Wang, S. 1985. Singing voice: Bright timbre, singer's formants and larynx position. In A. Askenfelt, S. Felicetti, E. Jansson & J. Sundberg (eds.) SMAC 83, Proceedings of the Stockholm Acoustics Conference, July 28–August 1. 1983, Vol. 1, 313-322. Stockholm.

Wedin, S., R. Leandersson & L. Wedin 1978. Evaluation of voice training. Folia Phoniatrica 30, 103-112.

Wendahl, R. 1963. Laryngeal analog synthesis of harsh voice quality. Folia Phoniatrica 15, 241-250.

Wendler, J., E. Doherty & H. Hollien 1980. Voice classification by means of long-term speech spectra. Folia Phoniatrica 32, 51-60.

Wendler, J., A. Rauhut & H. Krüger 1986. Classification of voice qualities. Journal of Phonetics 14, 483-488.

White, B. 1989. Singing techniques and vocal pedagogy. New York: Garland.

Williams, C. & K. Stevens 1972. Emotions and speech: Some acoustic correlates. Journal of the Acoustical Society of America 52, 1238-1250.

Yamazawa, H. & H. Hollien 1992. Speaking fundamental frequency of Japanese women. Phonetica 49, 128-140.

Yanagihara, N. 1967. Significance of harmonic changes and noise components in hoarseness. Journal of Speech and Hearing Research 10, 531-541.

APPENDIX

This study is a reinterpretation of issues, materials and results presented in the following publications. The numbers in virgules after each publication indicate to which chapters of the dissertation that publication particularly relates; in a more general sense they are related to the whole work.

1 Hurme, P. 1980a. Auto-monitored speech level and average speech spectrum. In P. Hurme (ed.) Ääni. puhe ja kieli: selosteita ja katsauksia, 119-127. Puheentutkimuksen alalta 2. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 19. /3, 4, 5.2/

2 Hurme, P. 1980b. Digitaalisen spektrianalyysin sovelluksia puheentutkimuksessa. [Applications of digital spectrum analysis in speech research.] In K. Suomi (ed.) Fonetiikan päivät – Oulu 1979, 40-49. Oulun yliopiston fonetiikan laitoksen julkaisuja 1. /3, 4, 5.1, 5.2/

3 Sonninen, A. & P. Hurme 1982. Clinical and acoustic observations of normal and hoarse voices. In P. Hurme (ed.) Normaalin ja patologisen puheen fonetiikkaa, 1-16. Puheentutkimuksen alalta 4. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 29. /3, 4, 5.1/

4 Hurme, P. & M. Pirinen 1984. Puheen keskiarvospektrien vaihtelusta. [Variation in long-term spectra of speech.] In U. Ikonen & T. Tikka (eds.) Fonetiikan päivät - Joensuu 1984, 23-36. Joensuun yliopisto. Kielitieteellisiä tutkimuksia 1. /3/

5 Sonninen, A. & P. Hurme 1984. Äänen laadun kuuntelukoe ja keskiarvospektrianalyysi: käheä ja normaali. [A listening test and long-term spectrum analyses of voice quality: hoarse and normal.] In U. Ikonen & T. Tikka (eds.) Fonetiikan päivät – Joensuu 1984, 76-92. Joensuun yliopisto. Kielitieteellisiä tutkimuksia 1. /3, 4, 5.1/

6 Hurme, P. & A. Sonninen 1985. Normal and disordered voice quality: Listening tests and long-term spectrum analyses. In P. Hurme (ed.) Puheentutkimuksen alalta 6, 49-72. Jyväskylän yliopiston viestintätieteiden laitoksen julkaisuja 1. /3, 4, 5.1/

7 Hurme, P. & A. Sonninen 1985. Acoustic correlates of pathological voices: Individuals with paralysis of recurrent nerve. In P. Hurme (ed.) Puheentutkimuksen alalta 6, 73-78. Jyväskylän yliopiston viestintätieteiden laitoksen julkaisuja 1. /3, 4, 5.1/

8 Hurme, P. 1985. Äänen laadun määrittelemisestä ja kuvaamisesta. [Definition and description of voice quality.] In O. Aaltonen & T. Hulkko (eds.) Fonetiikan päivät – Turku 1985, 65-84. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 26. /2/

9 Hurme, P. 1986. Äänen laadun arvioimisesta ja mittaamisesta. [Evaluation and measurement of voice quality.] In M. Lehtihalmes & A. Klippi (eds.) Logopedis-foniatrinen tutkimus Suomessa, 201-209. Suomen logopedis-foniatrisen yhdistyksen julkaisuja 19. Helsinki. /2, 3/

10 Hurme, P. & A. Sonninen 1986. Acoustic, perceptual and clinical studies of normal and dysphonic voice. Journal of Phonetics 14, 489-492. /3, 4, 5.1/

11 Sonninen, A. & P. Hurme 1992. On the terminology of voice research. Journal of Voice 6, 188-193. /2/

12 Sonninen, A., P. Hurme & J. Sundberg 1994. Physiological and acoustic observations of support in singing. In A. Friberg, J. Iwarsson, E. Jansson & J. Sundberg (eds.) SMAC 93, Proceedings of the Stockholm Music Acoustics Conference July 28 – August 1, 1993, 254-258. Publications issued by the Royal Swedish Academy of Music 79. Stockholm. /4, 5.3/

13 Sonninen, A. & P. Hurme 1996. Control of vocal strain in singing: Roentgenographic and acoustic observations. Manuscript submitted for publication. /4, 5.4/