State-of-the-Art Survey

LNAI 4343

Christian Müller (Ed.)

Speaker Classification I

Fundamentals, Features, and Methods

Springer

Christian Müller (Ed.)

Speaker Classification I

Fundamentals, Features, and Methods

Springer

E2007003354

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editor

Christian Müller
International Computer Science Institute
1947 Center Street, Berkeley, CA 94704, USA
E-mail: cmueller@icsi.berkeley.edu

Library of Congress Control Number: 2007932293

CR Subject Classification (1998): I.2.7, I.2.6, H.5.2, H.5, I.4-5

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-74186-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-74186-2 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12107810 06/3180 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 4343

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Preface

> "As well as conveying a message in words and sounds, the speech signal carries information about the speaker's own anatomy, physiology, linguistic experience and mental state. These speaker characteristics are found in speech at all levels of description: from the spectral information in the sounds to the choice of words and utterances themselves."

The best way to introduce this textbook is by using the words Volker Dellwo and his colleagues had chosen to begin their chapter "How Is Individuality Expressed in Voice?" While they use this statement to motivate the introductory chapter on speech production and the phonetic description of speech, it constitutes a framework of the entire book as well: What characteristics of the speaker become manifest in his or her voice and speaking behavior? Which of them can be inferred from analyzing the acoustic realizations? What can this information be used for? Which methods are the most suitable for diversified problems in this area of research? How should the quality of the results be evaluated?

Within the scope of this book the term *speaker classification* is defined as assigning a given speech sample to a particular class of speakers. These classes could be Women vs. Men, Children vs. Adults, Natives vs. Foreigners, etc. *Speaker recognition* is considered as being a sub-field of speaker classification in which the respective class has only one member (Speaker vs. Non-Speaker). Since in the engineering community this sub-field is explored in more depth than others covered by the book, many of the articles focus on speaker recognition. Nevertheless, the findings are discussed in the context of the broader notion of speaker classification where feasible.

The book is organized in two volumes. Volume I encompasses more general and overview-like articles which contribute to answering a subset of the questions above: Besides Dellwo and coworkers' introductory chapter, the "Fundamentals" part also includes a survey by David Hill, who addresses past and present speaker classification issues and outlines a potential future progression of the field.

The subsequent part is concerned with the multitude of candidate speaker "Characteristics." Tanja Schulz describes "why it is desirable to automatically derive particular speaker characteristics from speech" and focuses on language, accent, dialect, ideolect, and sociolect. Ulrike Gut investigates "how speakers can be classified into native and non-native speakers of a language on the basis of acoustic and perceptually relevant features in their speech" and compiles a list of the most salient acoustic properties of foreign accent. Susanne Schötz provides a survey about speaker age, covering the effects of ageing on the speech production mechanism, the human ability of perceiving speaker age, as well as its automatic recognition. John Hansen and Sanjay Patil "consider a range of issues associated with analysis, modeling, and recognition of speech under stress." Anton Batliner and Richard Huber address the problem of emotion classification focusing on the

specific phenomenon of irregular phonation or laryngealization and thereby point out the inherent problem of speaker-dependency, which relates the problems of speaker identification and emotion recognition with each other. The juristic implications of acquiring knowledge about the speaker on the basis of his or her speech in the context of emotion recognition is addressed by Erik Eriksson and his co-authors, discussing, "inter alia, assessment of emotion in others, witness credibility, forensic investigation, and training of law enforcement officers."

The "Applications" of speaker classification are addressed in the following part: Felix Burckhardt et al. outline scenarios from the area of telephone-based dialog systems. Michael Jessen provides an overview of practical tasks of speaker classification in forensic phonetics and acoustics covering dialect, foreign accent, sociolect, age, gender, and medical conditions. Joaquin Gonzalez-Rodriguez and Daniel Ramos point out an upcoming paradigm shift in the forensic field where the need for objective and standardized procedures is pushing forward the use of automatic speaker recognition methods. Finally, Judith Markowitz sheds some light on the role of speaker classification in the context of the deeper explored sub-fields of speaker recognition and speaker verification.

The next part is concerned with "Methods and Features" for speaker classification beginning with an introduction of the use of frame-based features by Stefan Schacht et al. Higher-level features, i.e., features that rely on either linguistic or long-range prosodic information for characterizing individual speakers are subsequently addressed by Liz Shriberg. Jacques Koreman and his co-authors introduce an approach for enhancing the between-speaker differences at the feature level by projecting the original frame-based feature space into a new feature space using multilayer perceptron networks. An overview of "the features, models, and classifiers derived from [...] the areas of speech science for speaker characterization, pattern recognition and engineering" is provided by Douglas Sturim et al., focusing on the example of modern automatic speaker recognition systems. Izhak Shafran addresses the problem of fusing multiple sources of information, examining in particular how acoustic and lexical information can be combined for affect recognition.

The final part of this volume covers contributions on the "Evaluation" of speaker classification systems. Alvin Martin reports on the last 10 years of speaker recognition evaluations organized by the National Institute for Standards and Technology (nist), discussing how this internationally recognized series of performance evaluations has developed over time as the technology itself has been improved, thereby pointing out the "key factors that have been studied for their effect on performance, including training and test durations, channel variability, and speaker variability." Finally, an evaluation measure which averages the detection performance over various application types is introduced by David van Leeuwen and Niko Brümmer, focusing on its practical applications.

Volume II compiles a number of selected self-contained papers on research projects in the field of speaker classification. The highlights include: Nobuaki Minematsu and Kyoko Sakuraba's report on applying a gender recognition system to estimate the "feminity" of a client's voice in the context of a voice

therapy of a "gender identity disorder"; a paper about the effort of studying emotion recognition on the basis of a "real-life" corpus from medical emergency call centers by Laurence Devillers and Laurence Vidrascu; Charl van Heerden and Etienne Barnard's presentation of a text-dependent speaker verification using features based on the temporal duration of context-dependent phonemes; Jerome Bellegarda's description of his approach on speaker classification which leverages the analysis of both speaker and verbal content information – as well as studies on accent identification by Emmanuel Ferragne and François Pellegrino, by Mark Huckvale and others.

February 2007 Christian Müller

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 4660: S. Džeroski, J. Todorovski (Eds.), Computational Discovery of Scientific Knowledge. X, 327 pages. 2007.

Vol. 4651: F. Azevedo, P. Barahona, F. Fages, F. Rossi (Eds.), Recent Advances in Constraints. VIII, 185 pages. 2007.

Vol. 4632: R. Alhajj, H. Gao, X. Li, J. Li, O.R. Zaïane (Eds.), Advanced Data Mining and Applications. XV, 634 pages. 2007.

Vol. 4626: R.O. Weber, M.M. Richter (Eds.), Case-Based Reasoning Research and Development. XIII, 534 pages. 2007.

Vol. 4617: V. Torra, Y. Narukawa, Y. Yoshida (Eds.), Modeling Decisions for Artificial Intelligence. XII, 502 pages. 2007.

Vol. 4612: I. Miguel, W. Ruml (Eds.), Abstraction, Reformulation, and Approximation. XI, 418 pages. 2007.

Vol. 4604: U. Priss, S. Polovina, R. Hill (Eds.), Conceptual Structures: Knowledge Architectures for Smart Applications. XII, 514 pages. 2007.

Vol. 4603: F. Pfenning (Ed.), Automated Deduction – CADE-21. XII, 522 pages. 2007.

Vol. 4597: P. Perner (Ed.), Advances in Data Mining. XI, 353 pages. 2007.

Vol. 4594: R. Bellazzi, A. Abu-Hanna, J. Hunter (Eds.), Artificial Intelligence in Medicine. XVI, 509 pages. 2007.

Vol. 4585: M. Kryszkiewicz, J.F. Peters, H. Rybinski, A. Skowron (Eds.), Rough Sets and Intelligent Systems Paradigms. XIX, 836 pages. 2007.

Vol. 4578: F. Masulli, S. Mitra, G. Pasi (Eds.), Applications of Fuzzy Sets Theory. XVIII, 693 pages. 2007.

Vol. 4573: M. Kauers, M. Kerber, R. Miner, W. Windsteiger (Eds.), Towards Mechanized Mathematical Assistants. XIII, 407 pages. 2007.

Vol. 4571: P. Perner (Ed.), Machine Learning and Data Mining in Pattern Recognition. XIV, 913 pages. 2007.

Vol. 4570: H.G. Okuno, M. Ali (Eds.), New Trends in Applied Artificial Intelligence. XXI, 1194 pages. 2007.

Vol. 4565: D.D. Schmorrow, L.M. Reeves (Eds.), Foundations of Augmented Cognition. XIX, 450 pages. 2007.

Vol. 4562: D. Harris (Ed.), Engineering Psychology and Cognitive Ergonomics. XXIII, 879 pages. 2007.

Vol. 4548: N. Olivetti (Ed.), Automated Reasoning with Analytic Tableaux and Related Methods. X, 245 pages. 2007.

Vol. 4539: N.H. Bshouty, C. Gentile (Eds.), Learning Theory. XII, 634 pages. 2007.

Vol. 4529: P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, W. Pedrycz (Eds.), Foundations of Fuzzy Logic and Soft Computing. XIX, 830 pages. 2007.

Vol. 4511: C. Conati, K. McCoy, G. Paliouras (Eds.), User Modeling 2007. XVI, 487 pages. 2007.

Vol. 4509: Z. Kobti, D. Wu (Eds.), Advances in Artificial Intelligence. XII, 552 pages. 2007.

Vol. 4496: N.T. Nguyen, A. Grzech, R.J. Howlett, L.C. Jain (Eds.), Agent and Multi-Agent Systems: Technologies and Applications. XXI, 1046 pages. 2007.

Vol. 4483: C. Baral, G. Brewka, J. Schlipf (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 327 pages. 2007.

Vol. 4482: A. An, J. Stefanowski, S. Ramanna, C.J. Butz, W. Pedrycz, G. Wang (Eds.), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. XIV, 585 pages. 2007.

Vol. 4481: J. Yao, P. Lingras, W.-Z. Wu, M. Szczuka, N.J. Cercone, D. Ślęzak (Eds.), Rough Sets and Knowledge Technology. XIV, 576 pages. 2007.

Vol. 4476: V. Gorodetsky, C. Zhang, V.A. Skormin, L. Cao (Eds.), Autonomous Intelligent Systems: Multi-Agents and Data Mining. XIII, 323 pages. 2007.

Vol. 4455: S. Muggleton, R. Otero, A. Tamaddoni-Nezhad (Eds.), Inductive Logic Programming. XII, 456 pages. 2007.

Vol. 4452: M. Fasli, O. Shehory (Eds.), Agent-Mediated Electronic Commerce. VIII, 249 pages. 2007.

Vol. 4451: T.S. Huang, A. Nijholt, M. Pantic, A. Pentland (Eds.), Artifical Intelligence for Human Computing. XVI, 359 pages. 2007.

Vol. 4438: L. Maicher, A. Sigel, L.M. Garshol (Eds.), Leveraging the Semantics of Topic Maps. X, 257 pages. 2007.

Vol. 4434: G. Lakemeyer, E. Sklar, D.G. Sorrenti, T. Takahashi (Eds.), RoboCup 2006: Robot Soccer World Cup X. XIII, 566 pages. 2007.

Vol. 4429: R. Lu, J.H. Siekmann, C. Ullrich (Eds.), Cognitive Systems. X, 161 pages. 2007.

Vol. 4428: S. Edelkamp, A. Lomuscio (Eds.), Model Checking and Artificial Intelligence. IX, 185 pages. 2007.

Vol. 4426: Z.-H. Zhou, H. Li, Q. Yang (Eds.), Advances in Knowledge Discovery and Data Mining. XXV, 1161 pages. 2007.

Vol. 4411: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 249 pages. 2007.

Vol. 4410: A. Branco (Ed.), Anaphora: Analysis, Algorithms and Applications. X, 191 pages. 2007.

Vol. 4399: T. Kovacs, X. Llorà, K. Takadama, P.L. Lanzi, W. Stolzmann, S.W. Wilson (Eds.), Learning Classifier Systems. XII, 345 pages. 2007.

Vol. 4390: S.O. Kuznetsov, S. Schmidt (Eds.), Formal Concept Analysis. X, 329 pages. 2007.

Vol. 4389: D. Weyns, H.V.D. Parunak, F. Michel (Eds.), Environments for Multi-Agent Systems III. X, 273 pages. 2007.

Vol. 4384: T. Washio, K. Satoh, H. Takeda, A. Inokuchi (Eds.), New Frontiers in Artificial Intelligence. IX, 401 pages. 2007.

Vol. 4371: K. Inoue, K. Satoh, F. Toni (Eds.), Computational Logic in Multi-Agent Systems. X, 315 pages. 2007.

Vol. 4369: M. Umeda, A. Wolf, O. Bartenstein, U. Geske, D. Seipel, O. Takata (Eds.), Declarative Programming for Knowledge Management. X, 229 pages. 2006.

Vol. 4343: C. Müller (Ed.), Speaker Classification I. X, 355 pages. 2007.

Vol. 4342: H. de Swart, E. Orłowska, G. Schmidt, M. Roubens (Eds.), Theory and Applications of Relational Structures as Knowledge Instruments II. X, 373 pages. 2006.

Vol. 4335: S.A. Brueckner, S. Hassas, M. Jelasity, D. Yamins (Eds.), Engineering Self-Organising Systems. XII, 212 pages. 2007.

Vol. 4334: B. Beckert, R. Hähnle, P.H. Schmitt (Eds.), Verification of Object-Oriented Software. XXIX, 658 pages. 2007.

Vol. 4333: U. Reimer, D. Karagiannis (Eds.), Practical Aspects of Knowledge Management. XII, 338 pages. 2006.

Vol. 4327: M. Baldoni, U. Endriss (Eds.), Declarative Agent Languages and Technologies IV. VIII, 257 pages. 2006.

Vol. 4314: C. Freksa, M. Kohlhase, K. Schill (Eds.), KI 2006: Advances in Artificial Intelligence. XII, 458 pages. 2007.

Vol. 4304: A. Sattar, B.-h. Kang (Eds.), AI 2006: Advances in Artificial Intelligence. XXVII, 1303 pages. 2006.

Vol. 4303: A. Hoffmann, B.-h. Kang, D. Richards, S. Tsumoto (Eds.), Advances in Knowledge Acquisition and Management. XI, 259 pages. 2006.

Vol. 4293: A. Gelbukh, C.A. Reyes-Garcia (Eds.), MICAI 2006: Advances in Artificial Intelligence. XXVIII, 1232 pages. 2006.

Vol. 4289: M. Ackermann, B. Berendt, M. Grobelnik, A. Hotho, D. Mladenič, G. Semeraro, M. Spiliopoulou, G. Stumme, V. Svátek, M. van Someren (Eds.), Semantics, Web and Mining. X, 197 pages. 2006.

Vol. 4285: Y. Matsumoto, R.W. Sproat, K.-F. Wong, M. Zhang (Eds.), Computer Processing of Oriental Languages. XVII, 544 pages. 2006.

Vol. 4274: Q. Huo, B. Ma, E.-S. Chng, H. Li (Eds.), Chinese Spoken Language Processing. XXIV, 805 pages. 2006.

Vol. 4265: L. Todorovski, N. Lavrač, K.P. Jantke (Eds.), Discovery Science. XIV, 384 pages. 2006.

Vol. 4264: J.L. Balcázar, P.M. Long, F. Stephan (Eds.), Algorithmic Learning Theory. XIII, 393 pages. 2006.

Vol. 4259: S. Greco, Y. Hata, S. Hirano, M. Inuiguchi, S. Miyamoto, H.S. Nguyen, R. Słowiński (Eds.), Rough Sets and Current Trends in Computing. XXII, 951 pages. 2006.

Vol. 4253: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. XXXII, 1301 pages. 2006.

Vol. 4252: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. XXXIII, 1335 pages. 2006.

Vol. 4251: B. Gabrys, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LXVI, 1297 pages. 2006.

Vol. 4248: S. Staab, V. Svátek (Eds.), Managing Knowledge in a World of Networks. XIV, 400 pages. 2006.

Vol. 4246: M. Hermann, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XIII, 588 pages. 2006.

Vol. 4223: L. Wang, L. Jiao, G. Shi, X. Li, J. Liu (Eds.), Fuzzy Systems and Knowledge Discovery. XXVIII, 1335 pages. 2006.

Vol. 4213: J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.), Knowledge Discovery in Databases: PKDD 2006. XXII, 660 pages. 2006.

Vol. 4212: J. Fürnkranz, T. Scheffer, M. Spiliopoulou (Eds.), Machine Learning: ECML 2006. XXIII, 851 pages. 2006.

Vol. 4211: P. Vogt, Y. Sugita, E. Tuci, C.L. Nehaniv (Eds.), Symbol Grounding and Beyond. VIII, 237 pages. 2006.

Vol. 4203: F. Esposito, Z.W. Raś, D. Malerba, G. Semeraro (Eds.), Foundations of Intelligent Systems. XVIII, 767 pages. 2006.

Vol. 4201: Y. Sakakibara, S. Kobayashi, K. Sato, T. Nishino, E. Tomita (Eds.), Grammatical Inference: Algorithms and Applications. XII, 359 pages. 2006.

Vol. 4200: I.F.C. Smith (Ed.), Intelligent Computing in Engineering and Architecture. XIII, 692 pages. 2006.

Vol. 4198: O. Nasraoui, O. Zaïane, M. Spiliopoulou, B. Mobasher, B. Masand, P.S. Yu (Eds.), Advances in Web Mining and Web Usage Analysis. IX, 177 pages. 2006.

Vol. 4196: K. Fischer, I.J. Timm, E. André, N. Zhong (Eds.), Multiagent System Technologies. X, 185 pages. 2006.

Vol. 4188: P. Sojka, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue. XV, 721 pages. 2006.

Vol. 4183: J. Euzenat, J. Domingue (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. XIII, 291 pages. 2006.

Vol. 4180: M. Kohlhase, OMDoc – An Open Markup Format for Mathematical Documents [version 1.2]. XIX, 428 pages. 2006.

Vol. 4177: R. Marín, E. Onaindía, A. Bugarín, J. Santos (Eds.), Current Topics in Artificial Intelligence. XV, 482 pages. 2006.

¥562.00元

Table of Contents

IV Methods and Features

V Evaluation

How Is Individuality Expressed in Voice? An Introduction to Speech Production and Description for Speaker Classification

Volker Dellwo, Mark Huckvale, and Michael Ashby

Department of Phonetics and Linguistics
University College London
Gower Street, London, WC1E 6BT
United Kingdom
v.dellwo@ucl.ac.uk, m.huckvale@ucl.ac.uk, m.ashby@ucl.ac.uk

Abstract. As well as conveying a message in words and sounds, the speech signal carries information about the speaker's own anatomy, physiology, linguistic experience and mental state. These speaker characteristics are found in speech at all levels of description: from the spectral information in the sounds to the choice of words and utterances themselves. This chapter presents an introduction to speech production and to the phonetic description of speech to facilitate discussion of how speech can be a carrier for speaker characteristics as well as a carrier for messages. The chapter presents an overview of the physical structures of the human vocal tract used in speech, it introduces the standard phonetic classification system for the description of spoken gestures and it presents a catalogue of the different ways in which individuality can be expressed through speech. The chapter ends with a brief description of some applications which require access to information about speaker characteristics in speech.

Keywords: Speech production, Phonetics, Taxonomy, IPA, Individuality, Speaker characteristics.

1 Introduction

Whenever someone speaks an utterance, they communicate not only a message made up of words and sentences which carry meaning, but also information about themselves as a person. Recordings of two people saying the same utterance will sound different because the process of speaking engages the neural, physiological, anatomical and physical systems of a specific individual in a particular circumstance. Since no two people are identical, differences in these systems lead to differences in their speech, even for the same message. The speaker-specific characteristics in the signal can provide information about the speaker's anatomy, physiology, linguistic experience and mental state. This information can sometimes be exploited by listeners and technological applications to describe and classify speakers, possibly allowing speakers to be categorised by age, gender, accent, language, emotion or

C. Müller (Ed.): Speaker Classification I, LNAI 4343, pp. 1–20, 2007.

health. In circumstances where the speaker is known to the listener, speaker characteristics may be sufficient to select or verify the speaker's identity. This leads to applications in security or forensics. The aim of this chapter is to provide a framework to facilitate discussion of these speaker characteristics: to describe ways in which the individuality of speakers can be expressed through their voices.

Always in the discussion of speaker characteristics, it must be borne in mind that a spoken utterance exists primarily for its communicative value – as an expression of a desire in the mind of the speaker to make changes in the mind of the listener. The study of the communicative value of utterances is the domain of Linguistics, which we take to include knowledge of articulation, phonology, grammar, meaning and language use. The study of speaker characteristics is in a sense parallel to this, where we concentrate on what a particular implementation of an utterance within the linguistic system tells us about the person speaking.

At first glance, it may appear that we should be able to separate speaker characteristics from message characteristics in a speech signal quite easily. There is a view that speaker characteristics are predominantly low level – related to the implementation in a particular physical system of a given set of phonetic gestures, while message characteristics operate at a more abstract level – related to the choice of phonetic gestures: the syllables, words and phrases that are used to communicate the meaning of a message. However this is to oversimplify the situation. Speakers are actually different at all levels, because speakers also differ in the way in which they realise the phonetic gestures, they vary in the inventory of gestures used, in the way in which gestures are modified by context, and in their frequency of use of gestures, words and message structure. A speaker's preferred means of morning greeting may help identify them just as much as their preferred average pitch.

To build a framework in which the many potential influences of an individual on his or her speech can be discussed, we have divided this chapter into three sections: section 2 provides an overview of vocal structures in humans, section 3 introduces the conventional principles of phonetic classification of speech sounds, while section 4 provides a discussion on how and on what levels speaker characteristics find their way into the speech signal and briefly discusses possible applications of this knowledge.

2 Vocal Apparatus

In this section we will give an overview of the physical structures in the human that are used in the physical generation of speech sounds. We will look at the anatomy of the structures, their movements and their function in speech. The first three sections look at the structures below the larynx, above the larynx and the larynx itself. The last section briefly introduces the standard signals and systems model of speech acoustics.

2.1 Sub-laryngeal Vocal Tract

Figure 1 shows the main anatomical structures that are involved in speaking. Looking below the larynx we see the lungs lying inside a sealed cavity inside the rib cage. The

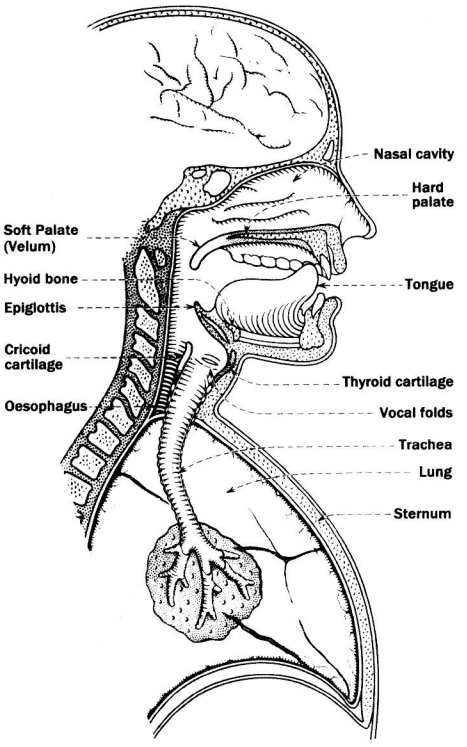

Fig. 1. Schematic diagram of the human organs of speech (Adapted from [1])

volume of the air spaces in the lungs can be varied from about 2 litres to about 6 litres in adults. The volume of the chest cavity and hence the volume of the lungs themselves is increased by lowering the diaphragm or raising the rib cage; the volume is decreased by raising the diaphragm or lowering the rib cage. The diaphragm is a dome of muscle, rising into the lower surface of the lungs, and tensing it causes it to flatten out and increase the size of the chest cavity; conversely relaxation of the diaphragm or action of the abdominal wall muscles makes the diaphragm more domed, reducing the size of the cavity. The external intercostal muscles bring the ribs closer together, but since they are pivoted on the vertebrae and are floating at the lower end of the rib cage, contraction of these muscles raises the rib cage and increases the volume of the chest cavity. The internal intercostal muscles can be used to depress the rib cage, and in combination with muscles of the abdominal wall, these can act to forcibly reduce the size of the chest cavity.

Changes in the size of the chest cavity affect the size of the lungs and hence the pressure of the air in the lung cavities. A reduction in pressure draws in air through the mouth or nose, through the pharynx, larynx and trachea into the lungs. A typical inspiratory breath for speech has a volume of about 1.5 litres, and is expended during speech at about 0.15 litres/sec [2]. One breath may be used to produce up to 30 seconds of speech. An increase in the pressure of air in the lungs forces air out

through the trachea, larynx, pharynx, mouth and nose. To produce phonation in the larynx, the lung pressure has to rise by at least 300Pa to achieve sufficient flow for vocal fold vibration. A more typical value is 1000Pa, that is 1% of atmospheric pressure.

Pressure is maintained during speech by a control mechanism that connects stretch receptors in the trachea, bronchioles and lung cavities to the muscles that control chest cavity volume. The stretch receptors provide information about the physical extension of the lung tissues which indirectly measures lung pressure. At large volumes the natural elasticity of the lungs would cause too high a pressure for speaking, so nerve activation on the diaphragm and external intercostal muscles is required to maintain a lower pressure, while at low volumes the elasticity is insufficient to maintain the pressure required for speaking, so nerve activation on the internal intercostal muscles and abdominal wall muscles is required to maintain a higher pressure.

2.2 The Larynx

The larynx is the major sound generation structure in speech. It sits in the air pathway between lungs and mouth, and divides the trachea from the pharynx. It is suspended from the hyoid bone which in turn is connected by muscles to the jaw, skull and sternum. This arrangement allows the larynx to change in vertical position. The larynx is structured around a number of cartilages: the cricoid cartilage is a ring that sits at the top of the trachea at the base of the larynx; the thyroid cartilage is a V shape with the rear legs articulating against the back of the cricoid cartilage and the pointed front sticking out at the front of the larynx and forming the "Adam's apple" in the neck; the two arytenoid cartilages sit on the cricoid cartilage at the back of the larynx.

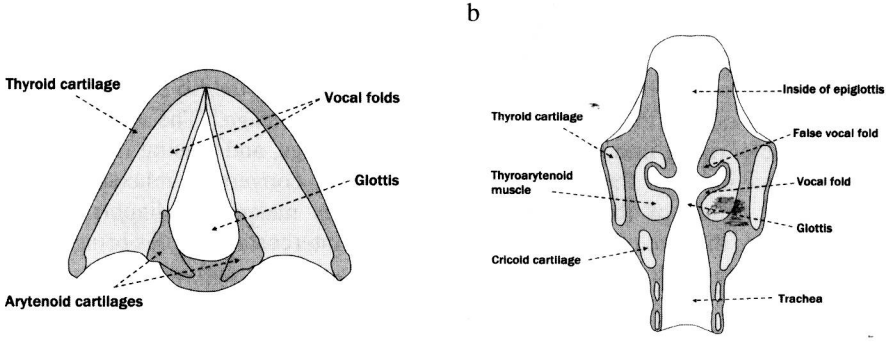

Fig. 2. Schematic diagrams of the larynx: (a) superior view, showing vocal folds, (b) vertical section, showing air passage

The vocal folds are paired muscular structures that run horizontally across the larynx, attached close together on the thyroid cartilage at the front, but connected at the rear to the moveable arytenoid cartilages, and forming an adjustable valve. For breathing the folds are held apart (abducted) at their rear ends and form a triangular opening known as the glottis. Alternatively, the arytenoids can be brought together

(adducted), pressing the folds into contact along their length. This closes the glottis and prevents the flow of air. If the folds are gently adducted, air under pressure from the lungs can cause the folds to vibrate as it escapes between them in a regular series of pulses, producing the regular tone called "voice". Abduction movements of the vocal folds are controlled by contraction of the posterior cricoarytenoid muscles, which cause the arytenoids to tilt and hence draw the rear of the vocal folds apart. Adduction movements are controlled by the transverse interarytenoid muscles and the oblique interarytenoid muscles which draw the arytenoids together, also the lateral cricoarytenoid muscles which cause the arytenoids themselves to swivel in such a way as to draw the rear of the folds together.

The open glottis position gives voiceless sounds, such as those symbolised [s] or [f]; closure produces a glottal stop, symbolised [ʔ], while voice is used for all ordinary vowels, and for many consonants. Commonly, consonants are in voiced-voiceless pairs; for example, [z] is the voiced counterpart of [s], and [v] the voiced counterpart of [f].

As well as adduction/abduction, the vocal folds can change in length and tension owing to movements of the thyroid and arytenoid cartilages and of changes to the muscles inside the vocal folds. These changes primarily affect the rate of vocal fold vibration when air is forced through a closed glottis. The cricothyroid muscles rock the thyroid cartilage down and hence stretch and lengthen the vocal folds. Swivelling of the arytenoid cartilages with the posterior and lateral interarytenoid muscles also moves the rear of the folds relative to the thyroid, and changes their length. Within the vocal folds themselves, the thyroarytenoid muscle can contract in opposition to the other muscles, and so increase the tension in the folds independently from their length.

Generally, changes in length, tension and degree of adduction of the vocal folds in combination with changes in sub-glottal pressure cause changes in the loudness, pitch and quality of the sound generated by phonation. Normal (modal) voice produces a clear, regular tone and is the default in all languages. In breathy voice (also called murmur), vibration is accompanied by audible breath noise. Other glottal adjustments include narrowing without vibration, which produces whisper, and strong adduction but low tension which produces an irregular, creaky phonation.

2.3 Supra-laryngeal Vocal Tract

Immediately above the larynx is the pharynx, which is bounded at the front by the epiglottis and the root of the tongue. Above the pharynx, the vocal tract branches into the oral and nasal cavities, see Fig. 1. The entrance to the nasal cavity is controlled by the soft palate (or velum) which can either be raised, to form a closure against the rear wall of the pharynx, or lowered, allowing flow into the nasal cavity and thus out of the nostrils. The raising of the soft palate is controlled by two sets of muscles: the tensor veli palatini and the levator veli palatini which enter the soft palate from above. Lowering of the soft palate is controlled by another two sets of muscles: the palatopharyngeus muscle and the palatoglossus muscle which connect the palate to the pharynx and to the back of the tongue respectively.

Air flowing into the oral cavity can eventually leave via the lip orifice, though its path can be controlled or stopped by suitable manoeuvres of the tongue and lips. The main articulators which change the shape and configuration of the supra-laryngeal vocal tract are the soft palate, the tongue, lips and jaw.

The upper surface of the oral cavity is formed by the hard palate, which is domed transversely and longitudinally, and is bordered by a ridge holding the teeth. In a mid-sagittal view, the portion of this behind the upper incisors is seen in section, and generally referred to as the alveolar ridge. The lower surface of the oral cavity consists of the tongue, a large muscular organ which fills most of the mouth volume when at rest. Various parts of the tongue can be made to approach or touch the upper surface of the mouth, and complete airtight closures are possible at a range of locations, the closure being made not only on the mid-line where it is usually visualised, but extending across the width of the cavity and back along the tongue rims. The position and shape of the tongue are controlled by two sets of muscles: the extrinsic muscle group lie outside the tongue itself and are involved in the protrusion of the tongue, the depression of the tip of the tongue, the forward-backward movement of the tongue and the raising and lowering of the lateral borders of the tongue. The intrinsic muscles lie within the body of the tongue and are involved in flattening and widening the tongue, lengthening and narrowing the tongue, and also raising and lowering the tongue tip. Together the many sets of muscles can move the bulk of the tongue within the oral cavity and change the shape of the remaining cavity, which in turn affects its acoustic properties.

The available space in the oral cavity and the distance between the upper and lower teeth can be altered by adjusting the jaw opening. Raising the jaw is performed mainly by the masseter muscle which connects the jaw to the skull, while lowering the jaw is performed by muscles that connect the jaw to the hyoid bone.

At the exit of the oral cavity, the lips have many adjustments that can affect the shape of the oral opening and even perform a complete closure. Lip movements fall into two broad categories: retrusive/protrusive movements largely performed by the orbicularis oris muscles that circle the lips, and lateral/vertical movements performed by a range of muscles in the cheeks that attach into the lips, called the muscles of facial expression.

2.4 Sound Generation

To a very good approximation, we can describe the generation of speech sounds in the vocal tract as consisting of two separate and independent processes. In the first process, a constriction of some kind in the larynx or oral cavity causes vibration and/or turbulence which gives rise to rapid pressure variations which propagate rapidly through the air as sound. In the second process, sound passing through the air cavities of the pharynx, nasal and oral cavities is modified in terms of its relative frequency content depending on the shape and size of those cavities. Thus the sound radiated from the lips and nostrils has properties arising from both the sound source and the subsequent filtering by the vocal tract tube. This approach is called the source-filter model of speech production.