State-of-the-Art Survey

LNAI 4441

Christian Müller (Ed.)

Speaker Classification II

Selected Projects

Springer

Christian Müller (Ed.)

Speaker Classification II

Selected Projects

Springer

E2007003358

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: 2007932403

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 3-540-74121-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-74121-3 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12104529 06/3180 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 4441

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 coworker's 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. 4441: C. Müller (Ed.), Speaker Classification. X, 309 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. 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.

¥484.00元

Table of Contents

A Study of Acoustic Correlates of Speaker Age

Susanne Schötz[1] and Christian Müller[2]

[1] Dept. of Phonetics, Centre for Languages and Literature,
Lund University, Sweden
susanne.schotz@ling.lu.se
[2] International Computer Science Institute, Berkeley, CA
cmueller@icsi.berkeley.edu

Abstract. Speaker age is a speaker characteristic which is always present in speech. Previous studies have found numerous acoustic features which correlate with speaker age. However, few attempts have been made to establish their relative importance. This study automatically extracted 161 acoustic features from six words produced by 527 speakers of both genders, and used normalised means to directly compare the features. Segment duration and sound pressure level (SPL) range were identified as the most important acoustic correlates of speaker age.

Keywords: Speaker age, Phonetics, Acoustic analysis, Acoustic correlates.

1 Introduction

Many acoustic features of speech undergo significant change with ageing. Earlier studies have found age-related variation in duration, fundamental frequency, SPL, voice quality and spectral energy distribution (both phonatory and resonance) [1,2,3,4,5,6]. Moreover, a general increase of variability and instability, for instance in F_0 and amplitude, has been observed with increasing age.

The purpose of the present acoustic study was to use mainly automatic methods to obtain normative data of a large number of acoustic features in order to learn how they are related to speaker age, and to compare the age-related variation in the different features. Specifically, the study would investigate features in isolated words, in stressed vowels, and in voiceless fricatives and plosives. The aim was to identify the most important acoustic correlates of speaker age.

2 Questions and Hypotheses

The research questions concerned acoustic feature variation with advancing speaker age: (1) What age-related differences in features can be identified in female and male speakers? and (2) Which are the most important correlates of speaker age?

C. Müller (Ed.): Speaker Classification II, LNAI 4441, pp. 1–9, 2007.

Based on the findings of earlier studies (cf. [5]), the following hypotheses were made: **Speech rate** will generally decrease with advancing age. **SPL range** will increase for both genders. $\mathbf{F_0}$ will display different patterns for female and male speakers. In females, F_0 will remain stable until around the age of 50 (menopause), when a drop occurs, followed by either an increase, decrease or no change. Male F_0 will decrease until around middle age, when an increase will follow until old age. **Jitter and shimmer** will either increase or remain stable in both women and men. **Spectral energy distribution** (spectral tilt) will generally change in some way. However, in the higher frequencies (spectral emphasis), there will be no change. **Spectral noise** will increase in women, and either increase or remain stable in men. **Resonance measures** in terms of formant frequencies will decrease in both female and male speakers.

3 Speech Material

The speech samples consisted of 810 female and 836 male versions of the six Swedish isolated words *käke* ['ɕɛ:kə] (jaw), *saker* ['sɑ:kəʁ] (things), *själen* ['ʃɛ:lən] (the soul), *sot* [su:t], *typ* [ty:p] (type (noun)) and *tack* [tak] (thanks). These words were selected because they had previously been used by the first author in a perceptual study [7] and because they contained phonemes which in a previous study had shown tendencies to contain age-related information (/p/, /t/, /k/, /s/, /ɕ/ and /ʃ/) [8]. The words were produced by 259 female and 268 male speakers, taken from the SweDia 2000 speech corpus [9] as well as from new recordings. All speakers were recorded using a Sony portable DAT recorder TCD-D8 and a Sony tie-pin type condenser microphone ECM-T140 at 48kHz/16 bit sampling frequency in a quiet home or office room. Figure 1 shows the age and gender distribution of the speakers.

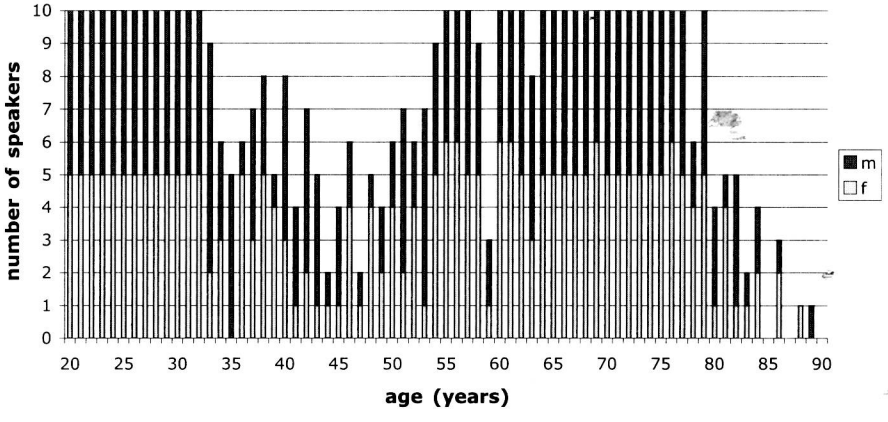

Fig. 1. Age distribution of the speakers used in this study

4 Method and Procedure

The acoustic analysis was carried out using mainly automatic methods. However, occasional manual elements were necessary in a few steps of the procedure. All words were normalised for SPL, aligned (i.e. transcribed into phoneme as well as plosive closure, VOT and aspiration segments) using several Praat [10] scripts and an automatic aligner[1]. Figure 2 shows an alignment example of the word *tack*. The alignments were checked several times using additional Praat scripts in order to detect and manually correct errors.

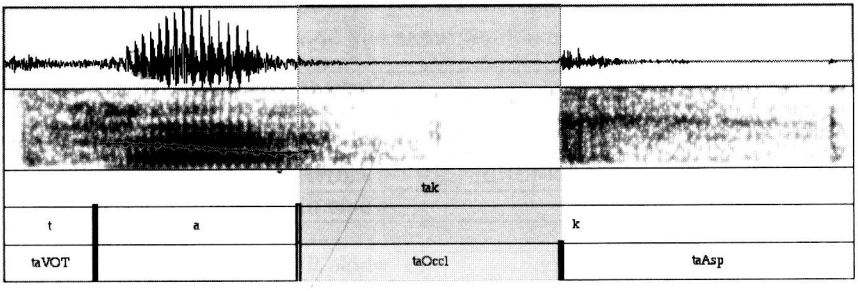

Fig. 2. Example of the word *tack*, aligned into word, phoneme, VOT, plosive closure and aspiration segments

The aligned words were concatenated; all the first word productions of a speaker were combined into one six-word sound file, all the second ones concatenated into a second file and so on until all words by all speakers had been concatenated. Figure 3 shows an example of an concatenated file.

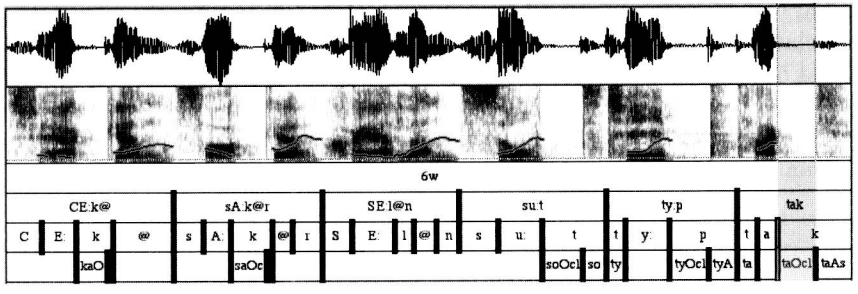

Fig. 3. Example of a concatenated file, aligned into word, phoneme, plosive closure, VOT and aspiration segments

[1] Originally developed by Johan Frid at the Department of Linguistics and Phonetics, Centre for Languages and Literature, Lund University, but further adapted and extended for this study by the first author.

A Praat script[1] extracted 161 acoustic features – divided into seven feature groups – from the concatenated words. Some features (e.g. syllables and phonemes per second, jitter and shimmer) were extracted only for all six words, while others (e.g. F_0, formant frequencies and segment duration) were extracted for several segments, including all six words and stressed vowels. Table 1 offers an overview of which segments were analysed in each feature group. Most features were extracted using the built-in functions in Praat. More detailed feature descriptions are given in [5].

Table 1. Segments analysed in each feature group (LTAS: long-term average spectra, HNR: harmonics-to-noise ratio, NHR: noise-to-harmonics ratio, sp.: spectral, str.: stressed)

Nr	*Feature group*	*Segments analysed*
1	syllables & phonemes per second	whole file
	segment duration (ms)	whole file, words, str. vowels,
2	sound pressure level (SPL) (dB)	fricatives, plosives (incl. VOT)
3	F_0 (Hz, semitones)	whole file, words, str. vowels
4	jitter, shimmer	
5	sp. tilt, sp. emphasis, inverse-filtered SB, LTAS	whole file
6	HNR, NHR, other voice measures	whole file, str. vowels
7	formant frequencies (F_1–F_5)	str. vowels
	sp. balance (SB)	fricatives and plosives

The analysis was performed with m3iCAT, a toolkit especially developed for corpus analysis [11]. It was used to calculate statistical measures, and to generate tables and diagrams, which displayed the variation of a certain feature as a function of age. The speakers were divided into eight overlapping "decade-based" age classes, based on the results (mean error ±8 years) of a previous human listening test [7]. There were 14 ages in each class (except for the youngest and oldest classes): 20, aged 20–27; 30, aged 23–37; 40, aged 33–47; 50, aged 43–57; 60, aged 53–67; 70, aged 63–77; 80, aged 73–87; 90, aged 83–89.

For each feature, m3iCAT calculated actual means (μ), standard deviations (σ) and normalised means ($\bar{\mu}$) for each age class. Normalisation involved mapping the domain of the values in the following way:

$$a_i = \frac{(v_i - mean)}{stdev} \tag{1}$$

where v_i represents the actual value, *mean* represents the mean value of the data and *stdev* represents the corresponding standard deviation. Occasionally, normalisations were also carried out separately for each gender. This was done in order to see the age-related variation more distinctly when there were large differences in the real mean values between female and male speakers, e.g. in F_0 and formant frequencies. Because of the normalisation process, almost all

values (except a few outliers) fall within the range between − 1 and + 1, which allows direct comparison of all features regardless of their original scaling and measurement units.

The values calculated for the eight age classes were displayed in tables, separately for female and male speakers. In addition, line graphs were generated for the age-class-related profiles or tendencies, with the age classes on the x-axis and the normalised mean values on the y-axis. The differences between the normalised mean values of all pairs of adjacent age classes are displayed as labels at the top of the diagrams (female labels above male ones). Statistical t-tests were carried out to calculate the significance of the differences; all differences except the ones within parentheses are statistically significant ($p \leq 0.01$). Figure 4 shows an example of a tendency diagram where the normalisations were carried out using all speakers (top), and the same tendencies but normalised separately for each gender (bottom).

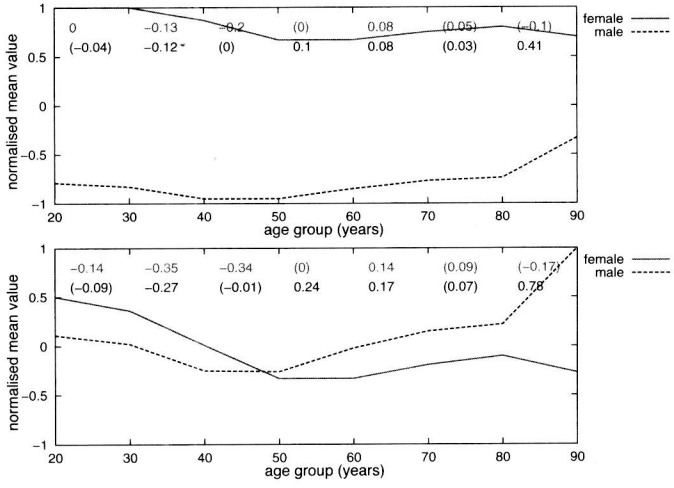

Fig. 4. Normalised tendencies for *mean F_0 (Hz) (all six words)*, 8 overlapping age classes, normalised for all speakers (top) and normalised separately for female and male speakers (bottom)

The advantage of using normalised means is that variation can be studied across features regardless of differences in the original scaling and units of the features. For instance, it allows direct comparison of the age-related variance between duration and F_0 by comparing the tendency for segment duration (in seconds) with the tendency for mean F_0 (in Hz).

5 Results

Due to the large number of features investigated, the results are presented by feature group (see Table 1). Moreover, only a few interesting results for each

feature group will be described, as it would be impossible to present the results for all features within the scope of this article. A more comprehensive presentation of the results is given in [5].

The number of syllables and phonemes per second generally decreased with increased age for both genders, while segment duration for most segments increased. The tendencies were less clear for the female than the male speakers. Figure 5 shows the results for all six words.

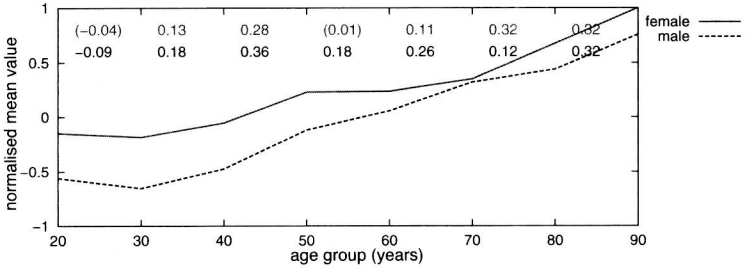

Fig. 5. Normalised tendencies for *duration (all six words)*

Average relative SPL generally either decreased slightly or remained constant with increased female and male age. The SPL range either increased or remained relatively stable with advancing age for both genders. Figure 6 shows the results for SPL range in the word *käke*. Similar tendencies were found for the other words, including the one without plosives; själen ['ʃɛːlən].

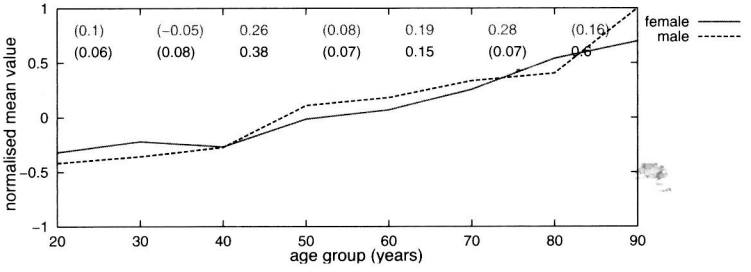

Fig. 6. Normalised tendencies for *SPL range (käke)*

Female F_0 decreased until age group 50 and then remained relatively stable. Male F_0 lowered slightly until age group 50, but then rose into old age. Due to the gender-related differences in F_0, the results for mean F_0 (Hz, all six words) are presented in Figure 7 as normalised separately for each gender to show clearer tendencies.