VOICE RECOGNIZER FOR PERSONAL IDENTIFICATION

VARAPORN PHOMVI-IN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2000

ISBN 974-663-997-8

3837393 SCCS/M : MAJOR : COMPUTER SCIENCE ; M.Sc. (COMPUTER SCIENCE)
KEY WORD : SPEAKER RECOGNITION / SPEAKER IDENTIFICATION
VARAPORN PHOMVI-IN : VOICE RECOGNIZER FOR PERSONAL IDENTIFICATION. THESIS ADVISOR : SUPACHAI TANGWONGSAN Ph.D., DAMRAS WONGSAWANG Ph.D. 65 p. ISBN 974-663-997-8

This research presents a model of a voice recognizer for personal identification. It is the process of automatically determining who the speaker is by matching his/her speech pattern to reference speakers in a database, which is obtained from a group of known speakers. The outcome of speaker identification is the one whose speech pattern is the best match to the known speech samples.

The voice recognizer system in this study consists of three major parts, namely: feature extraction, clustering, and identification. In the feature extraction, we initially detect voiced segments of speech samples by using ZCR (Short-Time Average Zero-Crossing Rate), then given as input to perform DFT (Discrete Fourier Transform) and compute the PSD (Power Spectrum Density). Next, the PSD vectors are used as training samples in a pattern-based training model and also used to calculate the distribution of the acoustic features by employing ML (Maximum Likelihood Procedure) to represent the statistical-based training model. In addition, gender classification was implemented in order to improve the performance during testing. This is based on the pitch frequency value that can be computed by using Auto-correlation function.

From the clustering method, we pre-detected the possible set of speakers who are acoustically similar to the test speaker by combining the scoring procedure and the Mahalanobis distance measure to optimize the distribution of each phonetic data. Finally, in the identification process, we employed the square-error pattern matching to determine the cohort set which is the last and smallest set of possible speakers. Then the decision algorithm is performed to obtain the most likely speaker from the cohort set.

The system is implemented in a stand-alone personal computer, in which Thai language is chosen as the word utterance for the evaluation. For noise-free speech recording environment, the voice recognizer is able to achieve the identification within accuracy up to 95%, while the identification error down to 5%. Concerning the computational time, the system is able to determine the speaker within 30 seconds, which is quite short in the sense of real environment. Moreover, the system requires minimal disk space because of using a small number of speech feature representatives, which significantly reduces the system pre-processing time.

3837393 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์ ; วท.ม. (วิทยาการคอมพิวเตอร์)

วราภรณ์ พรหมวิอินทร์ : ระบบตรวจรู้เสียงพูดเพื่อระบุตัวบุคคล (VOICE RECOGNIZER FOR PERSONAL IDENTIFICATION). คณะกรรมการควบคุมวิทยานิพนธ์ : ศุภชัย ตั้งวงศ์ศานต์, Ph.D., ดำรัส วงศ์สว่าง , Ph.D. 65 หน้า. ISBN 974-663-997-8

งานวิจัยนี้เป็นการนำเสนอการออกแบบต้นแบบของระบบตรวจรู้เสียงพูดเพื่อระบุตัวบุคคล คือกระบวนการที่ทำการบ่งชี้อย่างอัตโนมัติว่าใครคือผู้พูดโดยใช้การเทียบเคียงรูปแบบเสียงของผู้พูดทดสอบเข้ากับรูปแบบเสียงของผู้พูดอ้างอิงในฐานข้อมูลซึ่งเป็นการจัดเก็บรูปแบบเสียงจากกลุ่มผู้พูดที่ระบบรู้จัก ผลลัพธ์ที่ได้จากการตรวจรู้เสียงผู้พูดคือคนที่มีรูปแบบเสียงเข้ากันได้ดีที่สุดกับรูปแบบเสียงของผู้พูดทดสอบ

การศึกษาระบบตรวจรู้เสียงพูดนี้ประกอบด้วย 3 ส่วนหลักคือ การสกัดตัวแปรจำเพาะ (feature extraction) การแบ่งส่วน (clustering) และการบ่งชี้ (identification) ในการสกัดตัวแปรจำเพาะ ขั้นแรกจะตรวจจับเอาเฉพาะสัญญาณที่เป็นเสียงโดยใช้ ZCR (Short-time average zero-crossing rate) แล้วส่งสัญญาณที่ได้เข้า DFT (Discrete Fourier Transform) และทำการแปลงเป็น PSD (Power spectrum density) จากนั้นค่า PSD จะถูกนำไปเป็นตัวอย่างในแง่รูปแบบเสียงและถูกนำไปคำนวณหาค่าการกระจายของคุณลักษณะเสียงโดยใช้ ML (Maximum likelihood procedure) เพื่อเป็นตัวอย่างในแง่สถิติ นอกเหนือจากนี้เรายังได้พัฒนาการระบุเพศของผู้พูดโดยศึกษาการหาค่า pitch frequency value ที่สามารถคำนวณได้จากสมการ Auto-Correlation สำหรับการแบ่งส่วนใช้กลยุทธในการในการค้นหากลุ่มตัวอย่างของผู้พูดที่น่าจะมีรูปแบบเสียงคล้ายกับผู้พูดทดสอบ ขั้นตอนสุดท้ายเราจะใช้การเทียบเคียงรูปแบบเสียงของผู้พูดแต่ละคนในกลุ่มตัวอย่าง (cohort set) เข้ากับรูปแบบเสียงของผู้พูดทดสอบ และกระบวนการตัดสินใจจะทำการเลือกคนที่น่าจะใช่มากที่สุดออกมา

งานวิจัยนี้ได้พัฒนาบนเครื่องคอมพิวเตอร์ส่วนบุคคล โดยเลือกภาษาไทยเป็นคำที่ใช้ทดสอบ ในสภาพแวดล้อมที่เงียบปราศจากเสียงรบกวน ระบบตรวจรู้เสียงพูดให้ผลถูกต้องสูงถึง 95% ขณะที่ผลผิดพลาดจะอยู่ในราว 5% ในแง่ของเวลาในการคำนวณ ระบบสามารถระบุผู้พูดได้ภายใน 30 วินาที ซึ่งก็ถือว่าไม่มากเมื่อเทียบกับสถานะการณ์จริงของคน นอกจากนี้ระบบยังต้องการพื้นที่จัดเก็บฐานข้อมูลน้อย ซึ่งเป็นเหตุผลในแง่ประสิทธิภาพการคำนวณดีขึ้นอีกด้วย