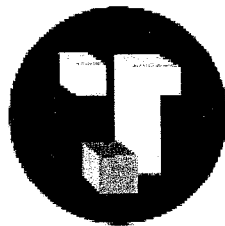


DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
DE TELECOMUNICACIÓN



Speed Up Strategies for the Creation of
Multimodal and Multilingual
Dialogue Systems

Autor

LUIS FERNANDO D'HARO

Ingeniero Electrónico

Director

RICARDO DE CÓRDOBA

Doctor Ingeniero de Telecomunicación

2009



Tribunal nombrado por el Sr. Rector Magfco. de la Universidad Politécnica de Madrid, el día...2...de...ABRIL...de 2009...

Presidente: JOSÉ MANUEL PARDO

Vocal: MICHAEL McTEAR

Vocal: RAMÓN LÓPEZ-ESTAR DELOADO

Vocal: DAVID MACÍAS GUARASA

Secretario: LUIS A. HERNÁNDEZ GÓMEZ

Suplente: VOLKER SCHLESS

Suplente: EMILIO SANGLIS ARNAL

Realizado el acto de defensa y lectura de la Tesis el día...13...de...MAYO...de 2009 en la E.T.S.I. de Telecomunicación.

Calificación: SOBRESALIENTE CUM LAUDE

EL PRESIDENTE

LOS VOCALES

Michael J. McTear

EL SECRETARIO

RESUMEN

Hoy por hoy, la mayoría de los sistemas comerciales y de investigación de atención telefónica se realizan mediante el uso de sofisticadas y completas plataformas que permiten especificar todos los detalles relacionados con el diseño, ejecución, y depuración de tales servicios. Pese a todas las funcionalidades y utilidades incluidas para acelerar el diseño y permitir servicios avanzados a los usuarios, la mayoría de ellas proponen el mismo tipo de aceleraciones y presentan limitaciones al desarrollo simultáneo del servicio para múltiples modalidades y perfiles de usuario.

En esta tesis se proponen diferentes estrategias de aceleración innovadoras, dinámicas e inteligentes que permiten predecir la información necesaria requerida para completar los diferentes aspectos del diseño, usando para ello información de la estructura del modelo de datos y del contenido de la base de datos del servicio, así como de la información acumulada a lo largo de todos los pasos ya realizados durante el diseño. Gracias a estas aceleraciones, la mayor parte del diseño del diálogo se reduce a confirmaciones por parte del diseñador de las "ofertas" que le hace la plataforma.

En concreto, se propone la generación semi-automática de diversos tipos de propuestas que pueden ser utilizadas para completar el flujo de la aplicación, las acciones que componen cada diálogo, o para solucionar problemas específicos de cada modalidad tales como la confirmación de datos al usuario y la presentación de las listas de resultados devueltos después de hacer una consulta a la base de datos del servicio. Así mismo, se propone la creación de diferentes asistentes que permiten acelerar la creación de las gramáticas usadas por el sistema de reconocimiento y la definición de las funciones de acceso a la base de datos. Los resultados obtenidos en sendas evaluaciones objetiva y subjetiva han permitido demostrar la viabilidad, relevancia y funcionalidad de estas aceleraciones y de la plataforma presentada.

Por otro parte, la amplia variedad de usuarios finales del servicio plantea diversos retos tales como la capacidad de identificar adecuadamente el idioma con el cual dirigirse a los usuarios, así como la posibilidad de proporcionar el servicio utilizando una u otra modalidad según las preferencias/necesidades de los usuarios o las condiciones actuales del diálogo.

En relación con las mejoras aplicadas al módulo de reconocimiento de idioma se ha implementado una nueva técnica para la incorporación de información contextual de más largo alcance en los modelos de lenguaje utilizados por el sistema basada en un ranking de n-gramas discriminativos. La técnica propuesta ha sido evaluada en la identificación de frases habladas en inglés y castellano obteniendo mejores tasas de reconocimiento que un sistema basado en PPRLM que usa modelos de lenguaje tradicionales gracias a la reducción del problema de falta de datos para el entrenamiento de los modelos de lenguaje de orden elevado lo que permite la utilización de modelos de mayor orden.

Finalmente, se han incorporado diversas mejoras a un módulo de traducción automática de voz a lengua de signos que permite ampliar las capacidades multimodales de la plataforma al permitir la prestación del mismo servicio, desarrollado con la plataforma de diálogo, a personas con discapacidad auditiva, permitiendo la traducción de los prompts del sistema en una secuencia animada reproducida por un avatar. En esta tesis se propone una técnica de adaptación innovadora que permite mejorar la calidad de las frases traducidas en situaciones en las que no hay suficientes datos para entrenar adecuadamente el modelo de lenguaje usado por el sistema de traducción. La adaptación se realiza a nivel de cuentas, mediante la técnica de Maximum-A-Posteriori (MAP), usando las cuentas de los n-gramas originales en el

idioma destino y las cuentas de ocurrencia de los n-gramas equivalentes en el idioma origen consultadas en la Web previamente y “traducidas” posteriormente a cuentas en el idioma destino usando un modelo de traducción basado en frases.

INDEX

ABSTRACT.....	I
RESUMEN.....	II
ACKNOWLEDGMENTS.....	IV
AGRADECIMIENTOS.....	V
INDEX.....	VI
INDEX OF FIGURES.....	X
INDEX OF TABLES.....	XIV
1 INTRODUCTION.....	1
1.1 Motivation.....	1
1.2 Objectives.....	3
1.2.1 Design platform.....	3
1.2.2 Language Identification System.....	3
1.2.3 Machine Translation.....	4
1.2.4 Relevant Definitions.....	4
1.3 Organization.....	5
2 STATE-OF-THE-ART.....	7
2.1 Development Platforms and Acceleration Strategies for Designing Multimodal Dialogue Systems.....	7 8
2.1.1 Commercial Platforms.....	16
2.1.2 Academic and Research Platforms.....	19
2.1.3 Research Platforms that Provide an Assisted Dialogue Design.....	24
2.1.4 Weaknesses of Commercial and Academic Platforms.....	24
2.2 Language Modelling.....	25
2.2.1 Statistical Language Models.....	35
2.2.2 Context-Free-Grammars (CFG's).....	36
2.3 Language Identification (LID).....	39
2.3.1 Description of the PPRLM Technique: Advantages and Disadvantages.....	41
2.4 Machine Translation.....	41
2.4.1 Current Approaches for Machine Translation.....	45
2.4.2 Word-based and Phrase-based Translation.....	47
2.4.3 Current Metrics for the Automatic Evaluation of Machine Translation Quality.....	50
2.4.4 Speech to Sign Language Translation.....	57
3 PLATFORM ARCHITECTURE.....	57
3.1 GDialogXML: Internal Descriptive Language for the Generated Models.....	58
3.2 FrameWork Layer.....	59
3.2.1 Application Description Assistant (ADA).....	60
3.2.2 Data Model Assistant (DMA).....	61
3.2.3 Data Connector Modelling Assistant (DCMA).....	62
3.3 Retrieval Layer.....	62
3.3.1 State Flow Modelling Assistant (SFMA).....	63
3.3.2 Retrieval Modelling Assistant (RMA).....	64
3.4 Dialogues Layer.....	64

3.4.1	User Modelling Assistant (UMA)	65
3.4.2	Modality Extension Retrieval Assistant for Speech (MERA-Speech)	65
3.4.3	Modality and Language Extension Assistant (MEA)	66
3.4.4	Dialogue Model Linker (DML)	69
3.4.5	Script Generators	69
3.4.6	Auxiliary Assistants	71
3.5	Runtime System	74
3.5.1	Speech Recognizer and Synthesizer	74
3.5.2	Animated Agent Used by the Sign Language Translation System	75
3.5.3	Distributed Platform and VoiceXML Interpreter (OpenVXI)	77
3.5.4	Portability and Use of Standards	82
3.6	Scope and Limitations	83
4	SPEED UP STRATEGIES APPLIED IN THE DIALOGUE DESIGN	85
4.1	Heuristics	86
4.2	Strategies Applied to the Data Model Assistant (DMA)	88
4.2.1	Semi-automatic Classes Proposals	89
4.2.2	Common Accelerations	90
4.3	Strategies Applied to the Data Connector Model Assistant (DCMA)	91
4.3.1	Definition of Relations between the Function Arguments and the Data Model	91
4.3.2	Automatic Generation of SQL Queries	93
4.4	Strategies Applied to the State Flow Model Assistant (SFMA)	94
4.4.1	Functionalities Included in the Graphical User Interface	94
4.4.2	Automatic State Proposals for Defining the Dialogue Flow	97
4.4.3	Automatic Unification of Slots for Mixed-Initiative Dialogues	99
4.5	Strategies Applied to the Retrieval Model Assistant (RMA)	101
4.5.1	Automatically Proposed Dialogues	101
4.5.2	Automatic Generation of Action Proposals in Each State	103
4.5.3	Automated Passing of Arguments between Actions	105
4.5.4	Mixed-Initiative and Over-Answering	106
4.5.5	Other Functionalities	109
4.6	Strategies Applied to the Modality Extension Retrieval Assistant for Speech (MERA-Speech)	109
4.6.1	Presentation of Object Lists	110
4.6.2	Confirmation Handling	113
4.7	Strategies Applied to Other Assistants	115
4.7.1	Modality and Language Extension Assistant (MEA)	116
4.8	Conclusions	122
5	EVALUATION OF THE ACCELERATION TECHNIQUES	125
5.1	Subjective Evaluation	125
5.1.1	Experimental setup	125
5.1.2	Evaluation results	127
5.2	Objective Evaluation	132
5.2.1	Experimental setup	132
5.2.2	Description of the evaluated tasks and results	135
5.2.3	Subjective survey	144
5.3	Conclusions	147
6	DEVELOPMENTS AND IMPROVEMENTS APPLIED TO THE RUNTIME SYSTEM	149

6.1	Language Identification System	150
6.1.1	<i>System Description</i>	150
6.1.2	<i>Proposed Technique: n-gram Frequency Ranking</i>	154
6.1.3	<i>Incorporation of additional information</i>	163
6.1.4	<i>Conclusions</i>	169
6.2	Automatic Translation of Dialogue Prompts into the Sign Language	170
6.2.1	<i>Runtime System for the Speech-to-Sign Language Translation System</i>	171
6.2.2	<i>Bilingual Corpus</i>	172
6.2.3	<i>Speech Recognition Results</i>	174
6.2.4	<i>Statistical Machine Translation System</i>	174
6.2.5	<i>Proposed Adaptation Technique</i>	177
6.2.6	<i>Language Model Experiments</i>	181
6.2.7	<i>Machine Translation Experiments</i>	181
6.2.8	<i>Conclusions</i>	183
7	CONCLUSIONS AND FUTURE WORK	185
7.1	CONCLUSIONS	185
7.1.1	<i>Dialogue Platform</i>	185
7.1.2	<i>LID System</i>	187
7.1.3	<i>Machine Translation System</i>	187
7.2	FUTURE WORK	188
7.2.1	<i>Dialogue Platform</i>	188
7.2.2	<i>LID System</i>	190
7.2.3	<i>Machine Translation System</i>	190
	BIBLIOGRAPHY	193
	APPENDIX A. LIST OF ABBREVIATIONS	209
	APPENDIX B. ADDITIONAL INFORMATION ABOUT CURRENT COMMERCIAL AND WEB-BASED PLATFORMS	211
B.1	Commercial Platforms	211
B.2	Web-Based Development Platforms	219
	APPENDIX C. TEMPLATES FOR THE CREATION OF AUTOMATIC DIALOGUES IN THE MERA-SPEECH ASSISTANT	225
C.1	Template for the Presentation of Lists of Objects	225
C.2	One Slot Confirmation	229
C.3	Mixed-Initiative Confirmation	230
C.4	One Slot Plus Over-Answering Confirmation	232
C.5	Mixed-Initiative Plus Over-Answering Confirmation	235
C.6	Simple Confirmation and Basic Dialogues	237
	APPENDIX D. QUESTIONNAIRE FOR EVALUATING THE APPLICATION GENERATION PLATFORM	239
D.1	Specific Questions by Assistant	239
D.1.1	<i>Questions regarding the assistant:</i>	239
D.2	General Questions about the AGP	240
D.2.1	<i>Advantages of using the AGP</i>	240
D.2.2	<i>Do you learn quickly how to make applications with the AGP?</i>	241
D.2.3	<i>How do you rate the overall appearance of the AGP (consistent, transparent, and intuitive)?</i>	242

<i>D.2.4 Do you find the various assistants of the AGP are well integrated?</i>	242
<i>D.2.5 Do you think non-experts could use the AGP efficiently?</i>	242
<i>D.2.6 Would you use this system in the future or recommend it to develop speech/Web applications?</i>	242
<i>D.2.7 If yes, how much would you be willing to pay for its use?</i>	242
APPENDIX E. DETAILED RESULTS OF THE OBJECTIVE EVALUATION OF THE PLATFORM	243