



University POLITEHNICA of Bucharest
**Faculty of Electronics, Telecommunications
and Information Technology**



Department of Electronic Devices, Circuits and Architectures
Speed – Speech and Dialogue Research Laboratory

Written and Spoken Language Processing Techniques with Applications in Automatic Speech Synthesis and Forensics

HABILITATION THESIS

– Abstract –

Author: Prof. Dragoş BURILEANU

Bucharest, July 2022

Abstract

This habilitation thesis describes the main coordinates of my academic and scientific career and presents several significant results obtained in my research activity after receiving the PhD title in April 1999, carried out within the *Speech and Dialogue* (Speed) Laboratory, Faculty of Electronics, Telecommunications and Information Technology, University “Politehnica” of Bucharest.

During over 30 years of activity, and especially in the last 23 years, I have mainly worked in the areas of digital signal processing, spoken language technologies, natural language processing, human-computer interaction, machine learning, biometric technologies, and forensic speech. This activity has been validated through an important number of scientific publications and through contributions in numerous research contracts and grants: I am the author or coauthor of 88 books, book chapters and journal / conference papers (72 after April 1999), 3 patents (one after 1999), and I was Project Manager or implementing expert for 24 national research projects (17 after 1999).

The thesis is structured in three parts. The first part presents several scientific results achieved in two main research directions: text-to-speech synthesis in Romanian and speech and audio forensics, respectively. It is worth mentioning that the described results are either entirely personal achievements or results to which my contribution has been substantial.

In the field of automatic speech synthesis, I continued (approximately during the years 1999 – 2012), the research activity started within the doctoral stage. The emphasis was especially on the development of the natural language processing part of a speech synthesis system based on concatenation of acoustic segments, capable to provide a high-quality synthesized speech (Chapter 1 of the thesis), as well as on the design and implementation of intelligent speech-enabled interfaces, both for embedded systems and for speech applications on communication channels (Chapter 2 of the thesis). This activity led to obtaining a number of results for the first time in Romania and also several new linguistic resources, publishing 3 book chapters (one in an international publishing house) and 25 scientific articles and invited papers, as well as to the completion of 7 national research projects (in 6 of them being Project Manager) and of a doctoral thesis that I coordinated. It should be mentioned that in the early 2000s, research in the field of automatic speech synthesis in Romanian was in its infancy, and free available linguistic resources (in electronic format) were almost non-existent; the current situation is much better.

My activity in the speech forensics domain started practically with the introduction of a new master program in the Faculty of Electronics and which I have been coordinating since 2011: “Multimedia Technologies in Biometrics and Information Security Applications” (BIOSINF). Given the present scale of the criminal phenomenon, the diversification of the methods and means used by the offenders, as well as the widespread introduction of audio-video recordings among the means of legal evidence, I introduced multimedia forensic expertise

as one of the three main directions of this study program. The major purpose was to train specialists in this field, capable of identifying information of interest in audio recordings, checking the authenticity of audio-video recordings, and performing forensic image processing, based on a good knowledge of the basic legal issues related to forensic examination systems and procedures. In this respect, my personal research interest was, on the one hand, on general speech and audio expertise, and, on the other hand (in recent years), on speech emotion recognition and stress and deceptive speech detection, using modern machine learning techniques, in the main framework of suspicious behavior monitoring in forensics and surveillance applications (Chapter 3 of the thesis). I must emphasize the idea that this last direction in the spoken language processing area is not only new in Romania but is also scarcely approached by the international scientific community. Finally, it is worth mentioning that this work has been validated, so far, by the publication (after 2011) of 19 scientific articles and invited papers, and the accomplishment of one national research project. In addition, two doctoral theses were completed, the third being in the final writing stage, theses coordinated by myself.

Chapter 4 concludes the first part with a brief description of several relevant scientific papers and a national research project that I coordinated (the project was carried out between 2007 and 2010).

The second part of the thesis is a review of the main professional, academic and research achievements. Chapter 5 contains a short presentation of my academic activity, and Chapter 6 describes the major coordinates of the scientific activity, a particular subject being the guidance of doctoral and master's students.

The last part of the thesis briefly discusses the future directions for my teaching and research activity, presenting both the possibilities to continue the previous work, as well as the new directions that I intend to approach in the following years.

Rezumat

Prezenta teză de abilitare descrie principalele coordonate ale carierei mele didactice și științifice și prezintă un număr de rezultate semnificative obținute în activitatea de cercetare după obținerea titlului de Doctor în Aprilie 1999, activitate desfășurată în cadrul *Laboratorului de tehnologia vorbirii și dialogul persoană – calculator (Speech and Dialogue – Speed)* din cadrul Facultății de Electronică, Telecomunicații și Tehnologia Informației, Universitatea POLITEHNICA din București.

În cei peste 30 de ani de activitate, și în special în ultimii 23 de ani, am lucrat mai ales în domeniile prelucrării digitale a semnalelor, tehnologiilor de prelucrare a limbajului vorbit, prelucrării limbajului natural, interacțiunii persoană – calculator, învățării automate, tehnologiilor biometrice și expertizei criminalistice a vorbirii. Această activitate a fost validată și s-a materializat într-un număr important de publicații științifice și de asemenea prin contribuțiile aduse în cadrul multor contracte și granturi de cercetare: sunt autor sau coautor pentru 88 de cărți, capitole de carte și articole publicate în jurnale sau volume ale unor conferințe de profil (dintre acestea, 72 au fost publicate după Aprilie 1999), 3 brevete de invenție (unul după 1999) și totodată am fost Director de Proiect sau participant în cadrul echipei de implementare pentru 24 proiecte de cercetare naționale (17 după 1999).

Teza este structurată în trei părți. Prima parte prezintă mai multe rezultate științifice obținute în cadrul a două mari direcții de cercetare: sinteza automată a vorbirii pornind de la text în limba română și respectiv expertiza criminalistică a vorbirii. Este important de subliniat faptul că rezultatele descrise sunt fie realizări personale, fie rezultate la care contribuția mea a fost substanțială.

În domeniul sintezei automate a vorbirii, am continuat (aproximativ în perioada anilor 1999 – 2012) activitatea de cercetare începută în timpul stagiului doctoral. Accentul a fost pus în mod special pe dezvoltarea părții de prelucrare a limbajului natural pentru un sistem de sinteză a vorbirii bazat pe concatenare de segmente acustice, capabil să genereze vorbire sintetizată de foarte bună calitate (Capitolul 1 al tezei), și, de asemenea, pe proiectarea și implementarea de interfețe inteligente cu răspuns prin voce, atât pentru sisteme portabile autonome, cât și pentru aplicații de transmitere a vorbirii pe canale de comunicație (Capitolul 2 al tezei). Această activitate a condus la obținerea unui număr de rezultate pentru prima oară în România și totodată a mai multor resurse lingvistice noi, ca și la publicarea a 3 capitole de carte (una într-o editură internațională) și a 25 de articole științifice și lucrări invitate; de asemenea, au fost finalizate 7 proiecte de cercetare naționale (pentru 6 dintre ele fiind Director de Proiect) și o teză de doctorat pe care am coordonat-o. În acest context, trebuie menționat faptul că la începutul anilor 2000, cercetarea în domeniul sintezei vorbirii în limba română era la început, resursele lingvistice disponibile (în format electronic) fiind aproape inexistente; situația este mult mai bună în prezent.

Activitatea mea în domeniul expertizei criminalistice a vorbirii a debutat practic o dată cu introducerea unui nou program de masterat în Facultatea de Electronică și pe care îl coordonez din anul 2011: “Tehnologii multimedia în aplicații de biometrie și securitatea informației” (BIOSINF). Având în vedere amploarea actuală a fenomenului infracțional, diversificarea metodelor și mijloacelor utilizate de infractori, precum și introducerea pe scară largă a înregistrărilor audio-video în rândul mijloacelor de probă judiciare, am introdus expertiza criminalistică în domeniul multimedia ca una dintre cele trei direcții principale ale acestui program de studii. Scopul major a fost formarea de specialiști în acest domeniu, capabili să identifice informații de interes în înregistrări audio, să verifice autenticitatea înregistrărilor audio-video și să efectueze expertize de imagini, având la bază și cunoștințe legate de aspectele legale ale expertizei criminalistice. În acest sens, activitatea mea de cercetare a fost îndreptată, pe de o parte, spre expertiza criminalistică a vocii și vorbirii, iar pe de altă parte (în ultimii ani), în direcția recunoașterii emoțiilor din vorbire și a detectării stresului și a minciunilor, folosind tehnici moderne de învățare automată, în cadrul mai larg al monitorizării comportamentului suspect în aplicațiile de criminalistică și supraveghere (Capitolul 3 al tezei). Doresc să accentuez faptul că această ultimă direcție din domeniul tehnologiilor vorbirii este nu numai nouă în România, dar totodată este foarte puțin abordată de către comunitatea științifică internațională. În sfârșit, menționez că această activitate a fost validată, până în prezent, prin publicarea (după 2011) a 19 articole științifice și lucrări invitate și realizarea unui proiect național de cercetare. În plus, au fost finalizate două teze de doctorat, a treia fiind în stadiul final de redactare, teze coordonate de mine.

Capitolul 4 încheie prima parte cu o scurtă descriere a mai multor lucrări științifice relevante și a unui proiect național de cercetare pe care l-am condus (proiect desfășurat între anii 2007 și 2010).

A doua parte a tezei trece în revistă principalele realizări profesionale, didactice și de cercetare. Capitolul 5 conține o scurtă prezentare a activității mele academice, iar Capitolul 6 descrie coordonatele majore ale activității științifice, un subiect special fiind îndrumarea doctoranzilor și a masteranzilor.

Ultima parte a tezei discută pe scurt direcțiile viitoare în activitatea de predare și cercetare, prezentând atât posibilitățile de continuare a cercetărilor anterioare, cât și noile direcții pe care intenționez să le abordez în anii următori.