

Nikos Fakotakis Andrey Ronzhin  
Rodmonga Potapova

# Speech and Computer

17th International Conference, SPECOM 2015  
Athens, Greece, September 20-24, 2015  
Proceedings Vol. II



University of Patras

## Volume Editors

Nikos Fakotakis  
University of Patras  
Electrical & Computer Engineering Department  
Patras, Greece  
E-mail: fakotaki@upatras.gr

Andrey Ronzhin  
St. Petersburg Institute for Informatics and Automation of  
the Russian Academy of Sciences  
Speech and Multimodal Interfaces Laboratory  
St. Petersburg, Russia  
E-mail: ronzhin@iiias.spb.su

Rodmonga Potapova  
Moscow State Linguistic University  
Institute of Applied  
and Mathematical Linguistics  
Moscow, Russia  
E-mail: rkpotapova@ya.ru

ISBN 978-960-530-160-6

University of Patras Press, Patras

# Table of Contents

## Conference Papers

Development of the Research Cloud Technology Stand-alone System (regarding Integrated SpeechDatabases) . . . . .	1
<i>Rodmonga Potapova, Vsevolod Potapov, and Irina Bazhenova</i>	
Prosodic Indicators of Russian Skype-Speech Semantic Content (Forensic Phonetic Aspect). . . . .	9
<i>Rodmonga Potapova and Vsevolod Potapov</i>	
Automatic Text Processing in Cross Language Computer Communication. . . . .	19
<i>Irina Borisova</i>	
Determination of the Internet Anonymity Influence on the Level of Aggression and Usage of Obscene Lexis . . . . .	29
<i>Rodmonga Potapova and Denis Gordeev</i>	
Statistical classification of small texts . . . . .	37
<i>Alexander A. Kharlamov, Alexey V. Maslov, and Rodmonga K. Potapova</i>	
Versatile Linguistic Database Annotation : Practical Issues and a New Flexible Approach. . . . .	45
<i>Rodmonga Potapova and Nikolay Bobrov</i>	
The Application of the WFST for the Problem of reconstruction of German Compound Words. . . . .	55
<i>Alexander Batalshchikov, Mikhail Kotov, Nikolay Shamraev, and Mikhail Zulkarneev</i>	
Russian Corpora for Speech Technology Applications: Purposes and Tools for Build-up . . . . .	63
<i>Anna Shirokova, Tatyana Platonova, Anna Smolina, Valeria Rogozhina, Mikhail Kotov and Anna Ovsyannikova</i>	

Gender Identification of Blog Authors: Do Men and Women Prefer Different Character Unigrams? .....	71
<i>Vasiliki Simaki, Athanasia Koumpouri, Iosif Mporas and Vasileios Megalooikonomou</i>	
Acoustic Differences between Standard Greek and Cypriot Greek .....	79
<i>Vasiliki Simaki, Iosif Mporas and Marianna Kondyli</i>	
Influence of a Content of Speech to Accuracy of Automatic Speaker Recognition .....	87
<i>Ivan Jokić, Stevan Jokić1, Vlado Delić and Zoran Perić</i>	
Simple Microphone Spill Reduction for On-Line Multi-Channel Meeting Speech Annotation .....	95
<i>Daniel Hladek, Jan Stas, and Jozef Juhar</i>	
Russian lexical units in the records of foreigners: a comprehensive electronic corpus .....	103
<i>Kamil Galiullin, Elena Gorobets, Guzel Karimullina, Rezeda Karimullina and Denis Martyanov</i>	
Foreign-Accented Speech Corpora: Annotation Challenge .....	111
<i>Viktoriya Zavyalova, Daria Poduzova, Diana Shaydullina and Ekaterina Lukina</i>	
<b>Author Index</b> .....	123

# Acoustic Differences between Standard Greek and Cypriot Greek

Vasiliki Simaki<sup>1, #</sup>, Iosif Mporas<sup>2, 3, \*</sup>, Marianna Kondyli<sup>1, \*</sup>

<sup>1</sup>Language and Literature Laboratory, Dept. of Educational Science and Early Childhood Education, University of Patras, 26500-Rion, Greece

<sup>2</sup>Intelligence Group, Wire Communications Laboratory, Dept. of Electrical and Computer Engineering, University of Patras, 26500-Rion, Greece

<sup>3</sup>Dept. of Computer and Informatics Engineering, TEI of Western Greece, 30020-Antirion, Greece

<sup>#</sup>simaki@ceid.upatras.gr

<sup>\*</sup>{imporas, kondyli}@upatras.gr

**Abstract.** The geographical linguistic variation is an important field of study in modern dialectology. The differences between the standard language and its dialectic variation can be observed in all levels of linguistic analysis. In this study we summarized researches in Phonetics about Cypriot and Standard Greek. We describe the differences in acoustic-level analysis between Cypriot and Standard Greek, detected in vowel and consonant system. This study attempts a synopsis of the differentiated acoustic features between the two varieties, and proposes eight rules that discriminate the Cypriot Greek from the Standard language. These rules enable experts in other disciplines than linguistics, i.e. in Speech Recognition, to use them in their research.

**Keywords:** acoustic features, Cypriot Greek, speech processing, Phonetics

## 1 Introduction

Ἀλλάξεν ὁ Μανωλιός τζι ἔβαλεν τα ρούχα του αλλιώς.

IPA Cypriot transcription: [ˈalaxenomanoliˈos ˈdʒevalenta ˈruχatˈaliˈos]

IPA Greek transcription: [ˈalaxeomanoliˈos ˈkevaleta ˈruχatuaˈliˈos]

An important field of study in modern dialectology is the geographical linguistic variation. The description of the geographical variation concerns the distribution of linguistic varieties in the space, the perspective of addressing a linguistic system to a specific location and the historical, social, political and cultural factors also implicated additionally to the local benchmark. The geographical coordinates of these varieties delimit the dialects and/or idioms. From a point of view, dialects are the facets of more spontaneous linguistic attitudes, which they do not incur the restrictions in evolution unlike the standard variations which appear to be more conservative. Concern-

ing the Modern Greek language, four major dialectic groups are recognized, according to most strict sense: the Tsakonian, the Southern Italian, the Pontic and the Capadocian. The Cretan and Cypriot for some researchers are fronted as idioms and for others as separate dialects [1]. In his recent study, which is quite extensive, Trudgill [2], reclaiming studies of Kontossopoulos [1] and Newton [3, 4], divides the map of Greece into fifteen Greek dialect-speaking areas according to phonological features.

Standard Greek is perceived as the standard variety in Greece after the abolition of diglossia [5]. The Cypriot Greek is the distinct local variety containing the systematic changes in all levels of linguistic analysis (phonology, grammar, semantics) as used in Cyprus today. The population of Standard Greek speakers is approximately 15 million (in Greece and abroad), and Cypriot speakers are approximately 1 million. The Cypriot Greek proves to be the first widespread variety of the Standard Greek, which is significantly different from the standard language and locally delimited. It is a vivid dialect, influenced by the English and the Standard Greek due to the language contact, and constitutes the exclusive choice of Cypriots apart of semi-formal and formal situations.

The distinction between the Standard and the Cypriot Greek is an important task, not only for dialectological research, but also for the speech and language technology area. An overview of the differences between the two varieties, standard language and dialect, may enable the automatic recognition of each variety in speech. The recognition of the dialectic vs the standard linguistic attitude is a challenging issue, and the detection of specific differentiated choices could be followed. The identification of distinctive features among the two varieties may lead to the transliteration of the dialectic speech into the standard language, in order to be comprehensible by all speakers. In spoken interaction applications, where speech technology engineers are interested in maximizing the recognition performance, the identification of the spoken dialect is essential for activating the corresponding acoustic model on the speech recognition engine [6], since there is not yet a developed system of orthographic standardization of Cypriot.

In this paper we present the differences between the Standard Modern Greek and the Cypriot Greek in phonetic level. We studied researches in phonetic analysis of Cypriot and Standard Greek and we generalized their findings. The variations in the vowel and the consonant systems, and generic rules of the distinctive features are proposed. We attempted to demonstrate that there exist phonemes in Cypriot Greek which are not used in Standard Greek Language and delimitate the two varieties in acoustic level.

The remainder of this paper is organized as follows. In Section 2 we describe the differences between the Standard Greek and the Cypriot Greek in phonetic-level. After a general presentation, we concentrate, in subsections 2.1 and 2.2, in the variation of the vowel and the consonant systems of Modern Greek and Cypriot respectively. In Section 3 we discuss our findings and future steps of our research.

## 2 Phonetic-level Differences Between Standard and Cypriot Greek

Differences between the Standard and the Cypriot Greek can be found in all levels of linguistic analysis. Several researches study linguistic phenomena in phonetic, phonological, morphological, syntactic, semantic and pragmatic level of analysis penetrating the Modern Greek and its dialectic utterances. Ralli [7] attempts a descriptive study of several syntactic and morphosyntactic phenomena applying to a range of Modern Greek dialects, in Cypriot among others, and makes theoretical analyses after the linguistic evidence. Theodorou et al. [8] deals with the development of relative clauses in Cypriot, its comprehension and production, and Arvaniti [9, 10], after an extensive research in Cypriot in many levels –phonetic, morphological, lexicon, etc.- investigates the existence of a Standard Cypriot language which contains the individual local accents around the island.

In this paper we concentrate in acoustic differences between the two varieties, the allophones related to the geographical idiom of Cyprus and the standard language. The Cypriot is classified as a southeastern idiom, sharing common accents with Rhodes, Karpathos, Kasos, Kastellorizo, Kos, Leros and Patmos. These islands have velar palatalization, geminates and final [n] retention. Another characteristic of these idioms is the south vocalism, namely, the vocal system remains undisturbed in non-stressed syllables. We present above the acoustic-level differences affecting the vowel system and the consonant system in separate sub-sections 2.1 and 2.2 respectively.

### 2.1 Vowel System

The Standard Greek has a typical five vowel system /i, e, a, o, u/. The Greek vowels do not exhibit much variation in terms of quality, and in casual speech, unstressed /i, u/ become devoiced and in some cases elided. The high /i/ is realized, depending on the position it occupies in the syllable [11].

In her study about the Cypriot vowel system, Eftychiou [12] examines the reduction of the high vowels [i] and [u] when a stop alveolar [t] is preceded at the end of an utterance, e.g. [em'fani'situ], [ˈfereti]. Eftichiou demonstrates how this reduction is measured in this context, and this reduction can range from the loss of higher formants, to fricated vowels and elided vowels, for which there was not found any visible evidence in the acoustic record. She also found that the reduction of the vowel appears to be in a complementary distribution with the reduction of the [t] which precedes, so that a fully realized [t] was more likely to be followed by a fully realized vowel. Tserdanelis [13] demonstrates that the [u] of the clitic [su] is not reduced when [su] is phrase-final and carries a 'continuation rise' in pitch, but it is reduced in phrase-initial position.

Acoustic studies of vowel quality comparing the Standard and the Cypriot Greek observe a faster speaking rate of Standard Greek and differences in quality between stressed and unstressed vowels [14, 15]. Differences in timing have also been reported, and in Standard Greek a faster speaking rate leads to a reduction in vowel dura-

tion, when in Cypriot the vowels are reduced similarly. It is observed that linguistic rhythm and vowel reduction interact, and it appears to be of great importance the investigation of segmental and prosodic environments, in order to reach useful conclusions about patterns of timing and prosody between the two varieties.

## 2.2 Consonant System

Cypriot displays morphophonetic archaisms in the consonant system. In Cypriot Greek, the nasal alveolar final /n/ is preserved as in Ancient Greek, and it is developed even in non-existent in Ancient Greek cases [1], e.g. [vu'non], ['lipin], ['polin] instead of [vu'no], ['lipi], ['poli] of Standard Greek. However, this phenomenon is not studied enough to extract a specific rule /or/ standardization.

The consonantal system of the Cypriot Greek is more complex when compared to Standard Greek, including post-alveolar, palatal consonants and a trill. The main difference in the consonantal system is the treatment of geminate consonants, between the Cypriot and the Standard Greek. In several phonetic studies [16, 17, 18, 19, 20], the geminates are examined in word-initial and intervocalic in many cases and level, with minor differences in their results. As a general conclusion, geminates are calculated 1.5/2 times longer than singletons. This durational difference depends on stress and the consonants' position in a word. In word-medial position there is a smaller difference in duration than word-initially. The difference is bigger though, when the geminate is followed by a stressed vowel.

Consonants /k, p, t/ are produced as aspirated [k<sup>h</sup>], [p<sup>h</sup>], [t<sup>h</sup>] accordingly, when found between vowels, e.g. ['kok<sup>h</sup>alon], [po't<sup>h</sup>e] instead of ['kokalo] and [po'te]. In Cypriot, the post-alveolar fricative [ʃ] is produced before [e] and [i] when an [χ] was in Standard Greek, e.g. ['χeri] → ['ʃerin]. Another variety is the alteration of stop palatal [s] and fricative alveolar [z] into [ʃ] and [ʒ] accordingly, when it is found before the semivowel [j], e.g. [ðja'kosja] → [ða'kosja] → [ðia'koʃa], [tra'pezja] → [tra'pezʒa]. Variation is also observed with the use of palatal fricative [j] in Cypriot Greek in place of [γ] following by [i, e] in Standard language.

A characteristic of the Cypriot, which appears also in the Cretan idiom, is the alteration of the stop velar consonant [k] to post alveolar [tʃ] when an [e] or [i] is following (tsitakism), e.g. [ke] → [tʃe], [e'ki] → [e'tʃi], [ke'ri] → [tʃe'ri]. At the same environment, when a nasal alveolar [n] is preceded, the sequence [ntʃ], becomes [dʃ], e.g. [tin ki'ra] → [tin tʃi'ra] → [tidʃi'ra]. The geminate consonants [f:, v:, p<sup>h</sup>:, tʃ:, z:, m:, n:, l:, s:, ʒ:, ɣ:, θ:, j:, ð:, k<sup>h</sup>:, t<sup>h</sup>:, ʃ:, ç<sup>h</sup>:, x:] reported to have more tense articulation than singletons after studies in the field [21, 22, 23]. Researchers have also examined the lateral consonants [24], and proved sensitive differences in duration and quality between the varieties, which cannot be treated as differential clues yet due to limited size of data.

In Table 1 we present examples of types of Standard Greek which are grouped according to the phonetic phenomenon, which differentiate the dialect from the standard language. These examples present the allophones of Cypriot containing phonemes which are not probatory in standard language.



**Table 1.** Examples of allophones, grouped according to the specific phonetic phenomenon

Rule #	Orthographic transcription of the Standard Type	Standard Greek transcription	Cypriot allophone transcription
1	εμφάνισή του (his appearance)	[em'fani'situ]	[em'fani'sit <sup>u</sup> ]
	φέρε τη (bring her)	[ˈfereti]	[ˈferent <sup>i</sup> ]
	δική του (his)	[di'kitu]	[di'kint <sup>u</sup> ]
2	βουνό (mountain)	[vu'no]	[vu'non]
	πόλη (city)	[ˈpoli]	[ˈpolin]
	έδωσε (gave)	[ˈedose]	[ˈedosen]
3	διάφορα (divers)	[ˈdiafora]	[ˈdjafora]
	κυάλια (spy glasses)	[ˈkiaʎia]	[ˈkjaʎia]
4	κόκκαλο (bone)	[ˈkokalo]	[ˈkok <sup>h</sup> alon]
	όπως (such)	[ˈopos]	[ˈop <sup>h</sup> os]
	ατομικός (individual)	[atomi'kos]	[at <sup>h</sup> omi'kos]
5	χέρι (hand)	[ˈçeri]	[ˈʃerin]
	χύμα (loose)	[ˈçima]	[ˈʃima]
6	διακόσια (two hundred)	[ðja'kosia]/[ðja'kosja]	[ðja'koʎa]
	κορίτσια (girls)	[ko'ritsia]/[ko'ritsja]	[ko'ritʎa]
	τραπέζια (tables)	[tra'pezia]/[tra'pezja]	[tra'peʎa]
7	εκεί (there)	[e'ki]	[e' tʃi]
	κυκεώνας (hotpot)	[kike'onas]	[tʃi tʃe'onas]
	καιρός (weather)	[ke'ros]	[tʃe'ros]
8	τον καιρό (the weather)	[toŋ ke'ro]	[to dʃe'ro]
	την κυρία (the lady)	[tiŋ ki'ria]	[ti dʃi'ria]
9	γέρος (old)	[ˈçeros]	[ˈʃeros]
	γιαγιά (grandmother)	[ɣia'ɣia]	[ʃia' ʃia]

In Table 2 we summarize the most important differences between the Standard and the Cypriot Greek in the form of rules. The rules are described and then transcribed after the International Phonetic Alphabet [25] and the SAMPA Computer Readable Phonetic Alphabet [26].

**Table 2.** The acoustic differences between the Standard and the Cypriot Greek

Rule #	Rule description	IPA transcription	SAMPA transcription
1	when toneless ending [ti] or [tu], the ending vowel [i, u] is reduced	[-tu, -ti] → [-t <sup>u</sup> , -t <sup>i</sup> ]	-tu, -ti→t <sup>u</sup> ,t <sup>i</sup>
2	final [n] is preservation (like in AG) and retention	[-n ]	[-n ]
3	[i] is replaced by semi-vowel [j] between a consonant and a vowel	c+[ i]+ v→c+[j]+v	c+i+v→c+j+v
4	[k,p,t] between vowels become aspirated [k <sup>h</sup> ,p <sup>h</sup> ,t <sup>h</sup> ]	v+[k,p,t]+v→ v+[k <sup>h</sup> ,p <sup>h</sup> ,t <sup>h</sup> ]+v	v+k,p,t+v→ v+k <sup>H</sup> ,p <sup>H</sup> ,t <sup>H</sup> +v
5	when [χ] precedes [e,i], becomes aspirated sibilant [ʃe,ʃi]	[χ]+[e,i] →[ʃe,ʃi]	X+e, i →Se, Si
6	when [s, z] precede the semi-vowel [j], the phoneme becomes aspirated [ʃ] and [ʒ] respectively	[s]+[j] →[ʃ] [z]+[j] →[ʒ]	s+j→S z+j→Z
7	when [k] precedes [e, i], becomes [tʃ]	[k]+[e,i] →[tʃe, tʃi]	k+e, i→tSe, tSi
8	when [n] precedes [tʃ], the phoneme becomes [dʃ]	[n, ŋ]+ [tʃ]→ [dʃ]	N+tS→dS
9	when [ɣ] is followed by [i, e], then becomes [ji, je]	[ɣi, ɣe] →[ ji, je]	Ge, Gi→jji, jje

The Cypriot allophones are not complementary about the Greek speakers. The Greek speakers do not use these allophones, and in many cases they are not even understood, when all standard language's utterances are comprehensible by Cypriots. An important element after our study is that the detection of Cypriot variants in speech precludes the occurrence of Greek speakers, and defines the variation used at the specific communicational situation.

### 3 Discussion and Conclusions

In this study we attempted the detection of the acoustic differences between the Standard Greek and the Cypriot Greek language. These characteristics identify the allophones between the two varieties, after linguistic studies, and they are described in an aggregated table in form of rules. This study was an effort to gather dialectological studies, empirical findings and theoretical analyses and 'transcribe' them into a form that facilitates computational studies in speech recognition. We aimed to demonstrate that the identification between these two varieties is possible in terms of allophones and differential acoustic features. These differences are listed, grouped

into generic phenomena and presented as phonetic rules discriminating the Cypriot from the Standard Greek.

Most studies in Cypriot as a Greek dialect reported common acoustic features, mostly affecting the consonant system. The major phenomena of allophones observed at this level are different enough to delimit the two varieties. Most studies demonstrate the sensitivity though of their findings, due to the small size of the data in some cases, and the minor deviation of the results and the differential attitudes in others. Owing to the challenging task of assorting the differentiated utterances, we proposed eight generic rules, describing the allophones of Cypriot utterances that are not part of the Standard Greek phonetic system.

## References

1. Kodosopoulos, N. (2000). *Dialects and Idioms of Modern Greek*. Ed. Grigori (In Greek).
2. Trudgill, P. (2003). Modern Greek dialects: A preliminary classification. *Journal of Greek linguistics*, 4(1), 45-63.
3. Newton, B. (1972). The generative interpretation of dialect: A study of Modern Greek phonology (Vol. 8). CUP Archive.
4. Newton, B. (1984). *Cypriot Greek revisited*. Linguistic Society of Athens.
5. Frangoudaki, A. (1992). Diglossia and the present language situation in Greece: A sociological approach to the interpretation of diglossia and some hypotheses on today's linguistic reality. *Language in Society*, 21(03), 365-381.
6. Mporas, I., Ganchev, T., & Fakotakis, N. (2008). Phonotactic recognition of greek and cypriot dialects from telephone speech. In *Artificial Intelligence: Theories, Models and Applications* (pp. 173-181). Springer Berlin Heidelberg.
7. Ralli, A. (2006). Syntactic and morphosyntactic phenomena in Modern Greek dialects: The state of the art. *Journal of Greek Linguistics*, 7(1), 121-159.
8. Theodorou, E., & Grohmann, K. K. (2012). The Acquisition of Relative Clauses in Cypriot Greek: Production and Comprehension. *Revista Diacritica*, 26(1), 270-299.
9. Arvaniti, A. (2006). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *Department of Linguistics, UCSD*.
10. Arvaniti, A. (1999). Standard Modern Greek. *Journal of the International Phonetic Association*, 29(02), 167-172.
11. Arvaniti, A. (2007). Greek Phonetics. *Journal of Greek Linguistics*, 8, 97-208.
12. Eftychiou, E. (2007). Stop-vowel coarticulation in Cypriot Greek. In *Proceedings of the International Congress of Phonetic Sciences XVI*, pp. 517-520.
13. Tserdanelis, G. (2003). Phonetic variation in syntactically ambiguous sentences. In *Studies in Greek Linguistics*, 23, pp. 931-942.
14. Loukina, A. (2009). Phonetic Variation in Spontaneous Speech: Vowel and Consonant Reduction in Modern Greek Dialects. *Oxford University Working Papers in Linguistics, Philology & Phonetics*, 36.
15. Arvaniti, A. (2010). A (brief) review of Cypriot phonetics and phonology. Voskos A., Goutsos D. & Moser A. éds, *The Greek Language in Cyprus from Antiquity to Today*, Athènes, Université d'Athènes, 107-124.
16. Arvaniti, A. and Tserdanelis, G. (2000). On the phonetics of geminates: evidence from Cypriot Greek. In *Proceedings of International Conference on Spoken Language Processing (ICSLP2000)*, Beijing.

17. Tserdanelis, G. and Arvaniti, A. (2001). The acoustic characteristics of geminate consonants in Cypriot Greek. In *Proceedings of the Fourth International Conference on Greek Linguistics*, pp. 29-36. Thessaloniki: University Studio Press.
18. Payne, E. and Eftychiou, E. (2006). Prosodic shaping of consonant germination in Cypriot Greek. In *Phonetica*, 63, pp. 175-198.
19. Christodoulou, C. (2007). On the phonetics of Cypriot-Greek geminate plosives. In *Journal of the Acoustical Society of America*, 121:3169.
20. Muller, J. (2002). On the theoretical implications of Cypriot Greek initial geminates. In *Journal of Greek Linguistics*, 3, pp. 115-137.
21. Arvaniti, A. (2001). Cypriot Greek and the phonetics and phonology of geminates. In *Proceedings of the First International Conference of Modern Greek Dialects and Linguistic Theory* (pp. 19-29).
22. Arvaniti, A. (1999). Effects of speaking rate on the timing of single and geminate sonorants. In *Proceedings of the XIVth International Congress of Phonetic Sciences* (Vol. 1, pp. 599-602).
23. Arvaniti, A. (1998). Phrase accents revisited: comparative evidence from standard and Cypriot greek. In *ICSLP*.
24. Loukina, A. (2009, June). Towards the acoustic analysis of lateral consonants in Modern Greek dialects: a preliminary study. In *Online Proceedings of the Forth International Conference on Modern Greek Dialects and Linguistic Theory, Chios* (pp. 120-132).
25. International Phonetic Association (Ed.). (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
26. Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. *Handbook of standards and resources for spoken language systems*, 4.