

**ΠΡΟΣ**

- 1) Όλα τα μέλη ΔΕΠ του Τμήματος Επιστήμης Υπολογιστών
- 2) Τους εκπροσώπους των Μεταπτυχιακών φοιτητών του Τμήματος Επιστήμης Υπολογιστών
- 3) Την Επταμελή Εξεταστική Επιτροπή
- 4) Όλα τα μέλη της Πανεπιστημιακής Κοινότητας

**Πρόσκληση σε Δημόσια Παρουσίαση της Διδακτορικής Διατριβής του**

**κ. Καφεντζή Γεωργίου**

Την Παρασκευή, 20 Ιουνίου 2014 και ώρα 12:00 στην αίθουσα Τηλεδιάσκεψης Κ206 του Πανεπιστημίου Κρήτης στο Ηράκλειο, θα γίνει η δημόσια παρουσίαση και υποστήριξη της Διδακτορικής Διατριβής του υποψηφίου διδάκτορος του Τμήματος Επιστήμης Υπολογιστών κ. Καφεντζή με θέμα:

**“ Προσαρμόσιμα Ημιτονοειδή Μοντέλα για Σήματα Φωνής με εφαρμογές σε Μετασχηματισμούς Φωνής και Ανάλυση Σημάτων Ήχου ”**

**“Adaptive Sinusoidal Models for Speech with Applications in Speech Modifications and Audio Analysis”**

**ΠΕΡΙΛΗΨΗ**

Η Ημιτονοειδής Μοντελοποίηση είναι μια από τις πιο ευρέως χρησιμοποιούμενες παραμετρικές μεθόδους για την επεξεργασία σήματος φωνής και ήχου. Η ακριβής εκτίμηση των ημιτονοειδών παραμέτρων (πλάτη, συχνότητες, και φάσεις) είναι ένα κρίσιμο σημείο για τη ακριβή αναπαράσταση των σημάτων που αναλύονται. Στην παρούσα εργασία, με βάση τις πρόσφατες εξελίξεις στην ημιτονοειδή ανάλυση, προτείνουμε υψηλής ανάλυσης, προσαρμόσιμα ημιτονοειδή μοντέλα για συστήματα

ανάλυσης, σύνθεσης, και τροποποίησης ομιλίας. Στόχος μας είναι να προσφέρουμε συστήματα που αναπαριστούν σήματα φωνής με εξαιρετικά ακριβή και συμπαγή τρόπο.

Εμπνευσμένοι από πρόσφατα προταθέντα μοντέλα, όπως το προσαρμόσιμο Σχεδόν - Αρμονικό Μοντέλο (aQHM) και το προσαρμόσιμο Αρμονικό Μοντέλο (aHM), διατυπώνουμε τη θεωρία της προσαρμόσιμης Ημιτονοειδούς Μοντελοποίησης και προτείνουμε ένα μοντέλο που ονομάζεται εκτεταμένο προσαρμόσιμο Σχεδόν - Αρμονικό Μοντέλο (eaQHM), το οποίο είναι ένα μη παραμετρικό μοντέλο, ικανό να προσαρμόσει τα στιγμιαία πλάτη και φάσεις των συναρτήσεων βάσης του στα τοπικά χρονικά μεταβαλλόμενα χαρακτηριστικά του σήματος της φωνής, αμβλύνοντας έτσι τη γνωστή υπόθεση της τοπικής στασιμότητας. Αποδεικνύεται ότι το eaQHM παρουσιάζει υψηλότερες επιδόσεις από το aQHM στην ανάλυση και ανασύνθεση των έμφωνων τμημάτων φωνής. Με βάση το eaQHM, ένα υβριδικό σύστημα ανάλυσης / σύνθεσης ομιλίας παρουσιάζεται (eaQHNM), μαζί με μια υβριδική έκδοση του του aHM (aHNM). Επιπλέον, παρουσιάζουμε κίνητρα για μια αναπαράσταση του σήματος της φωνής σε όλο το φάσμα και σε όλη τη διάρκεια του, χρησιμοποιώντας το eaQHM, αναπαριστώντας έτσι όλα τα μέρη του σήματος της φωνής, με υψηλής ανάλυσης AM-FM ημίτονα. Η αξιολόγηση δείχνει ότι η προσαρμοσιμότητα και η σχεδόν-αρμονικότητα είναι αρκετή για να παράξει πολύ υψηλή ποιότητα στην ανασύνθεση των άφωνων τμημάτων της φωνής. Στη συνέχεια, παρουσιάζεται το σύστημα πλήρους φάσματος ανάλυσης και σύνθεσης βασισμένο στο eaQHM, το οποίο υπερτερεί συστημάτων που θεωρούνται state-of-the-art, υβριδικά ή πλήρους ανάλυσης, στην ανάλυση και ανασύνθεση φωνής. Η υπεροχή του στην ποιότητα ανασύνθεσης επιβεβαιώθηκε με αντικειμενικές και υποκειμενικές αξιολογήσεις.

Όσον αφορά τις εφαρμογές, το eaQHM και το aHM εφαρμόζονται σε μετασχηματισμούς φωνής (κλιμάκωση χρόνου και κλιμάκωση θεμελιώδους συχνότητας). Οι μετασχηματισμοί που προκύπτουν είναι υψηλής ποιότητας, ακολουθώντας πολύ απλούς κανόνες, σε σύγκριση με άλλα συστήματα state-of-the-art. Οι έννοιες της σχετικής φάσης και της καθυστέρησης σχετικής φάσης είναι ζωτικής σημασίας για την ανάπτυξη μετασχηματισμένου σήματος με χαρακτηριστικά αναλλοίωτου σχήματος, χωρίς τεχνικά ελαττώματα, και υψηλής ποιότητας. Τα αποτελέσματα δείχνουν ότι τα συστήματα βασισμένα στην αρμονικότητα προτιμούνται έναντι αυτών της σχεδόν-αρμονικότητας, λόγω της απλότητας της αναπρασάσης. Επιπλέον, το eaQHM εφαρμόζεται στο πρόβλημα της μοντελοποίησης σημάτων ήχου, και συγκεκριμένα ήχων μουσικών οργάνων. Το eaQHM αξιολογείται και συγκρίνεται με state-of-the-art συστήματα, και έχει υψηλές επιδόσεις όσον αφορά την ποιότητα επανασύνθεσης, αναπαριστώντας με επιτυχία τα στάδια της επίθεσης, της μετάβασης, και της στατικότητας ενός ήχου μουσικού οργάνου. Τέλος, μια άλλη προτεινόμενη εφαρμογή έγκειται στην ανάλυση και ταξινόμηση της εκφραστικής ομιλίας. Το eaQHM εφαρμόζεται στην ανάλυση της εκφραστικής ομιλίας, παρέχοντας τις στιγμιαίες παραμέτρους του ως χαρακτηριστικά που μπορούν να χρησιμοποιηθούν στην αναγνώριση και ταξινόμηση, βασισμένη σε διανυσματικούς κβαντιστές, εκφραστικής ομιλίας. Αν και τα ημιτονοειδή μοντέλα δεν έχουν χρησιμοποιηθεί αρκετά σε τέτοιες εφαρμογές, τα αποτελέσματα είναι ελπιδοφόρα.

## ABSTRACT

Sinusoidal Modeling is one of the most widely used parametric methods for speech and audio signal processing. The accurate estimation of sinusoidal parameters (amplitudes, frequencies, and phases) is a critical task for close representation of the analyzed signal. In this thesis, based on recent advances in sinusoidal analysis, we propose high resolution adaptive sinusoidal models for analysis, synthesis, and modifications systems of speech. Our goal is to provide systems that represent speech in a highly accurate and compact way.

Inspired by the recently introduced adaptive Quasi-Harmonic Model (aQHM) and adaptive Harmonic Model (aHM), we overview the theory of adaptive Sinusoidal Modeling and we propose a model named the extended adaptive Quasi-Harmonic Model (eaQHM), which is a non-parametric model able to adjust the instantaneous amplitudes and phases of its basis functions to the underlying time-varying characteristics of the speech signal, thus significantly alleviating the so-called local stationarity hypothesis. The eaQHM is shown to outperform aQHM in analysis and resynthesis of voiced speech. Based on the eaQHM, a hybrid analysis/synthesis system of speech is presented (eaQHNM), along with a hybrid version of the aHM (aHNM). Moreover, we present motivation for a full-band representation of speech using the eaQHM, that is, representing all parts of speech as high resolution AM-FM sinusoids. Experiments show that adaptation and quasi-harmonicity is sufficient to provide transparent quality in unvoiced speech resynthesis. The full-band eaQHM analysis and synthesis system is presented next, which outperforms state-of-the-art systems, hybrid or full-band, in speech reconstruction, providing transparent quality confirmed by objective and subjective evaluations.

Regarding applications, the eaQHM and the aHM are applied on speech modifications (time and pitch scaling). The resulting modifications are of high quality, and follow very simple rules, compared to other state-of-the-art modification systems. The concepts of relative phase and relative phase delays are crucial for the development of artefact-free, shape-invariant, high quality modifications. Results show that harmonicity is preferred over quasi-harmonicity in speech modifications due to the embedded simplicity of representation. Moreover, the full-band eaQHM is applied on the problem of modeling audio signals, and specifically of musical instrument sounds. The eaQHM is evaluated and compared to state-of-the-art systems, and is shown to outperform them in terms of resynthesis quality, successfully representing the attack, transient, and stationary part of a musical instrument sound. Finally, another application is suggested, namely the analysis and classification of emotional speech.

The eaQHM is applied on the analysis of emotional speech, providing its instantaneous parameters as features that can be used in recognition and Vector-Quantization-based classification of the emotional content of speech. Although the sinusoidal models are not often used in such tasks, results are promising.

Supervisor: Professor Yannis Stylianou