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 της

κας. Κουτσογιαννάκη Μαρίας

Την Τετάρτη, 16 Δεκεμβρίου 2015 και ώρα 11:00 στην αίθουσα Τηλεδιάσκεψης Κ206 του Τμήματος Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης στο Ηράκλειο, θα γίνει η δημόσια παρουσίαση και υποστήριξη της Διδακτορικής Διατριβής της υποψήφιας διδάκτορας του Τμήματος Επιστήμης Υπολογιστών κας. Κουτσογιαννάκη Μαρίας με θέμα:

"Αύξηση της καταληπτότητας της ομιλίας χρησιμοποιώντας ιδιότητες καταληπτής ομιλίας"

"*Intelligibility enhancement of Casual speech based on Clear speech properties*"

ΠΕΡΙΛΗΨΗ

Όταν ένας άνθρωπος επικοινωνεί με έναν συνάνθρωπό του, προσαρμόζει αντανακλαστικά την ομιλία του ανάλογα με το περιβάλλον στο οποίο βρίσκεται αυτός (π.χ. παρουσία θορύβου) ή ο συνομιλητής του (π.χ. βαρήκοος), παράγοντας διαφορετικά στυλ ομιλίας (Καθαρή ομιλία, ομιλία Lombard) σε σχέση με το αν η επικοινωνία του ήταν ανεμπόδιστη (Πρόχειρη ομιλία). Τα στυλ αυτά ομιλίας διαφέρουν ανάλογα με το είδος του εμποδίου στο επικοινωνιακό κανάλι ή/και ανάλογα με τον ομιλητή. Παρουσιάζουν όμως ένα κοινό χαρακτηριστικό: την αυξημένη καταληπτότητα. Η ανάπτυξη αλγορίθμων που εκμεταλλεύονται τα ακουστικά χαρακτηριστικά τέτοιων στυλ ομιλίας θα μπορούσε να είναι επωφελής

στην Τεχνολογία Φωνής. Πολλές τεχνολογικές εφαρμογές αναζητούν μεθόδους βελτιστοποίησης της καταληπτότητας των συσκευών που παράγουν συνθετική φωνή. Πέρα από την εμπορική εκμετάλλευση των εφαρμογών αυτών (κινητά τηλέφωνα, συστήματα πλοήγησης, συστήματα τηλεφωνικής υποστήριξης πελατών), πολύ σημαντική είναι η συνεισφορά τους στον ιατρικό τομέα ως βοηθητικά μέσα επικοινωνίας ατόμων με προβλήματα ομιλίας και ακοής. Ωστόσο, η τρέχουσα τεχνολογία φωνής είναι «κωφή», δεν μπορεί δηλαδή να προσαρμοστεί στα δυναμικώς μεταβαλλόμενα περιβάλλοντα ούτε στην ιδιαιτερότητα του ακροατή, όπως ο άνθρωπος.

Η εργασία αυτή προτείνει την ανάπτυξη αλγορίθμων που «μιμούνται» τον τρόπο ανθρώπινης ομιλίας σε δύσκολες συνθήκες επικοινωνίας, συνεισφέροντας στην ανάπτυξη έξυπνων τεχνολογικών συστημάτων φωνής. Συγκεκριμένα, το στυλ ομιλίας του οποίου τα χαρακτηριστικά αναλύονται και χρησιμοποιούνται για την αύξηση της καταληπτότητας της Πρόχειρης ομιλίας είναι η Καθαρή ομιλία. Σε αντίθεση με άλλα στυλ ομιλίας, η Καθαρή ομιλία είναι καταληπτή από διάφορους ακροατές (ομόγλωσσους και μη, με προβλήματα ακοής, με κοχλιακά εμφυτεύματα, ηλικιωμένους, με μαθησιακές δυσκολίες κλπ) σε διάφορες συνθήκες περιβάλλοντος (με ή χωρίς θόρυβο, σε περιβάλλοντα αντήχησης).

Ένα σημαντικό μέρος της εργασίας αυτής αναλύει και συγκρίνει τα χαρακτηριστικά της Πρόχειρης και Καθαρής ομιλίας. Από την σύγκριση αυτή, αναδεικνύονται διαφορές κυρίως στην προσωδία, στον φωνηεντικό χώρο, στην φασματική ενέργεια και στο πλάτος διαμόρφωσης της χρονικής περιβάλλουσας του σήματος φωνής. Βασιζόμενοι στις μετρίσιμες αυτές διαφορές, προτείνουμε μετασχηματισμούς που βελτιώνουν την καταληπτότητα του σήματος Πρόχειρης ομιλίας. Σε σύγκριση με state-of-the-art συστήματα μετασχηματισμού, οι δικές μας τεχνικές (1) είναι χαμηλής υπολογιστικής πολυπλοκότητας (2) μπορούν να εφαρμοστούν ανεξαρτήτως ομιλητή ή σήματος (3) διατηρούν την ποιότητα του αρχικού σήματος (4) εφαρμόζονται άμεσα χωρίς την ανάγκη εκπαίδευσης των δεδομένων και προϋπαρξης του σήματος Καθαρής φωνής.

Οι προτεινόμενοι αλγόριθμοι αξιολογηθήκαν ως προς την καταληπτότητα και την ποιότητα με αντικειμενικές μετρικές καταληπτότητας και με υποκειμενικά ακουστικά τέστ από ομόγλωσσους και αλλόγλωσσους ακροατές χωρίς την ύπαρξη θορύβου, μέσα σε θορυβώδη περιβάλλοντα και σε περιβάλλοντα αντήχησης. Η αξιολόγηση δείχνει ότι οι μετασχηματισμοί που προτείνουμε αυξάνουν την καταληπτότητα της πρόχειρης ομιλίας τόσο σε περιβάλλοντα θορύβου όσο και σε περιβάλλοντα αντήχησης για ομόγλωσσους και αλλόγλωσσους ακροατές. Συγκεκριμένα, η τεχνική φασματικού μετασχηματισμού, επονομαζόμενη ως Mix-filtering, αυξάνει την καταληπτότητα του σήματος ομιλίας σε περιβάλλοντα θορύβου και αντήχησης ενώ διατηρεί την ποιότητα του σήματος, εν αντιθέσει με άλλους αλγορίθμους. Επιπλέον, η προτεινόμενη τεχνική αύξησης του πλάτους των διαμορφώσεων της χρονικής περιβάλλουσας, αναφερθείσα ως DMod, αυξάνει την καταληπτότητα της Πρόχειρης ομιλίας κατά 30% σε περιβάλλοντα θορύβου. Ο αλγόριθμος DMod, εμπνέεται όχι μόνο από χαρακτηριστικά της Καθαρής ομιλίας αλλά και από μη γραμμικές λειτουργίες που λαμβάνουν χώρα στην βασική μεμβράνη του ανθρώπινου κοχλία.

Επιτυγχάνει δε, πέρα από την αύξηση της καταληπτότητας, την εισαγωγή μιας νέας μεθόδου χειρισμού των διαμορφώσεων της περιβάλλουσας του σήματος. Τα αποτελέσματα της μελέτης αυτής δείχνουν την ύπαρξη μιας σύνδεσης ανάμεσα στις διαμορφώσεις της χρονικής περιβάλλουσας και στον τρόπο αντίληψης και επεξεργασίας του ήχου από την βασική μεμβράνη του ανθρώπινου κοχλία, ανοίγοντας τον δρόμο για την ανάλυση και κατανόηση της ομιλίας ως κύματα διαμορφώσεων.

Επόπτης Διδακτορικής Διατριβής: Καθηγητής Γιάννης Στυλιανού

ABSTRACT

In adverse listening conditions (e.g. presence of noise, hearing-impaired listener etc.) people adjust their speech in order to overcome the communication difficulty and successfully deliver their message. This remarkable adjustment produces different speaking styles compared to unobstructed speech (casual speech) that vary among speakers and conditions, but share a common characteristic; high intelligibility. Developing algorithms that exploit acoustic features of intelligible human speech could be beneficial for speech technology applications that seek methods to enhance the intelligibility of "speaking-devices". Besides the commercial orientation (e.g., mobile telephone, GPS, customer service systems) of these applications, most important is their medical context, providing assistive communication to people with speech or hearing deficits. However, current speech technology is deaf, meaning that it cannot adjust, like humans do, to the dynamically changing real environments or to the listener's specificity.

This work proposes signal modifications based on the acoustic properties of a high intelligible human speaking style, the clear speech, assisting in the development of smart speech technology systems that "mimic" the way people produce intelligible speech. Unlike other speaking styles, clear speech has a high intelligibility impact on various listening populations (native and non-native listeners, hearing impaired, cochlear implant users, elderly people, people with learning disabilities etc.) in many listening conditions (quiet, noise, reverberation).

A significant part of this work is devoted to the comparative analysis between casual and clear speech, which reveals differences on prosody, vowel spaces, spectral energy and modulation depth of the temporal envelopes. Based on these observed and measured differences between the two speaking styles, we propose modifications for enhancing the intelligibility of casual speech. Compared to other state-of-the-art modification systems, our modification techniques (1) do not require excessive computation (2) are speaker and speech independent (3) maintain speech quality (4) are explicit, since they do not require statistical training and the preexistence of clear speech recordings.

Evaluations on intelligibility and quality are performed objectively using recently proposed objective intelligibility scores and subjectively with listening tests conducted by native and non native listeners in noisy environments (speech shaped noise, SSN), reverberation and in quiet. Results show that our modifications enhance speech intelligibility in SSN and reverberation for native and non-native listeners. Specifically, the proposed spectral modification technique, namely Mix-filtering, increases the intelligibility of speech in noise and reverberation while maintains the quality of the original signal, unlike other intelligibility boosters. Moreover, a modulation depth enhancement technique called DMod, increases speech intelligibility more than 30% in SSN. DMod algorithm is inspired by both clear speech properties and by the non-linear phenomena that take place in the basilar membrane. DMod not only achieves to enhance speech intelligibility, but it introduces a novel method for manipulating the modulation spectrum of the signal. Results of this study indicate a connection of the modulations of the temporal envelopes with speech perception and specifically with processes that take place on the basilar membrane of human ear and pave the way for analyzing and comprehending speech in terms of modulations.

Supervisor: Professor Yannis Stylianou