

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ
Εφαρμοσμένα μαθηματικά για μηχανικούς

Φθινόπωρο 2015
Γ. Τζιρίτας, Καθηγητής

Εργασία Μαθήματος

**Διάκριση μουσικής/φωνής βασισμένη στο πλάτος του σήματος και στο ρυθμό
εναλλαγής προσήμου**

Η εργασία έχει στόχο να εξάγει από σήματα ήχου χαρακτηριστικά χρήσιμα για την τμηματοποίησή τους και για την ταξινόμηση στη συνέχεια των τμημάτων που προκύπτουν. Οι μετρήσεις θα γίνουν τοπικά με χρήση ενός γραμμικού συστήματος με χρονική απόκριση άπειρης διάρκειας

$$H(z) = \frac{(1 - \alpha)^2}{(1 - \alpha z^{-1})(1 - \alpha z)},$$

που υλοποιείται με χρήση δύο αναδρομικών σχέσεων. Το πρώτο χαρακτηριστικό είναι το πλάτος του σήματος $x(n)$ που δίδεται από τη σχέση

$$A(n) = \sqrt{\sum_{m=-\infty}^{\infty} h(m)x^2(n-m)}.$$

Το δεύτερο χαρακτηριστικό είναι ο ρυθμός εναλλαγής προσήμου που δίδεται ως εξής

$$Z(n) = \sum_{m=-\infty}^{\infty} h(m)|\text{sign}(x(n-m)) - \text{sign}(x(n-m-1))|.$$

1. Να υπολογισθούν το πλάτος και ο ρυθμός εναλλαγής προσήμου, με προτεινόμενη τιμή $\alpha = 0.95$, και να παρασταθούν γραφικά.
2. Να χωρισθεί το σήμα, συνολικής διάρκειας σε πλήθος δειγμάτων N , σε μη επικαλυπτόμενα πλαίσια διάρκειας $T = 20$ ms, οπότε κάθε πλαίσιο θα έχει $B = 0.02F_s$ δείγματα, όπου F_s είναι η συχνότητα δειγματοληψίας. Να ευρεθεί σε κάθε πλαίσιο η μέση τιμή του πλάτους

$$A_b(k) = \frac{1}{B} \sum_{n=1}^B A((k-1)B + n), k = 1, \dots, \frac{N}{B}$$

και η μέση τιμή του ρυθμού εναλλαγής προσήμου

$$Z_b(k) = \frac{1}{B} \sum_{n=1}^B Z((k-1)B + n), k = 1, \dots, \frac{N}{B}.$$

3. Να φιλτραρισθεί το σήμα $A_b(k)$ με το ακόλουθο σύστημα

$$H(z) = \frac{(1 - \alpha)^2(z - z^{-1})}{(1 - \alpha z)(1 - \alpha z^{-1})}$$

που υπολογίζει διαφορές και να ληφθεί η απόλυτη τιμή της απόκρισης, έστω $D_b(k)$.

4. Να εντοπισθούν τα τοπικά μέγιστα του $D_b(k)$ που υπερβαίνουν το διπλάσιο της μέσης του τιμής και απέχουν τουλάχιστον 200 ms. Οι στιγμές των μεγίστων είναι πιθανές στιγμές αλλαγής από φωνή σε μουσική ή αντίστροφα.
5. Για να ληφθεί απόφαση για τα τμήματα που προέκυψαν από την ανίχνευση των τοπικών μεγίστων, υπολογίζεται εμπειρικά για κάθε τμήμα η πιθανότητα το γινόμενο $A_b(k)Z_b(k)$ να είναι μικρότερο από $2400C/F_s$. Η σταθερά C υπολογίζεται ως η μέγιστη τιμή του $A_b(k)$ για το εξεταζόμενο τμήμα και για $Z_b(k) < 2400/F_s$. Αν η πιθανότητα είναι μεγαλύτερη από 0.75, τότε αποφασίζεται ότι πρόκειται για σήμα φωνής, αλλιώς μουσικής.

Χρήσιμες συναρτήσεις : audioread, diff, filtfilt, findpeaks, mean.

Σήμα : TestSegment.wav