

ΠΑΡΑΜΟΡΦΩΜΕΝΗ ΟΜΙΛΙΑ

ΕΙΣΑΓΩΓΗ

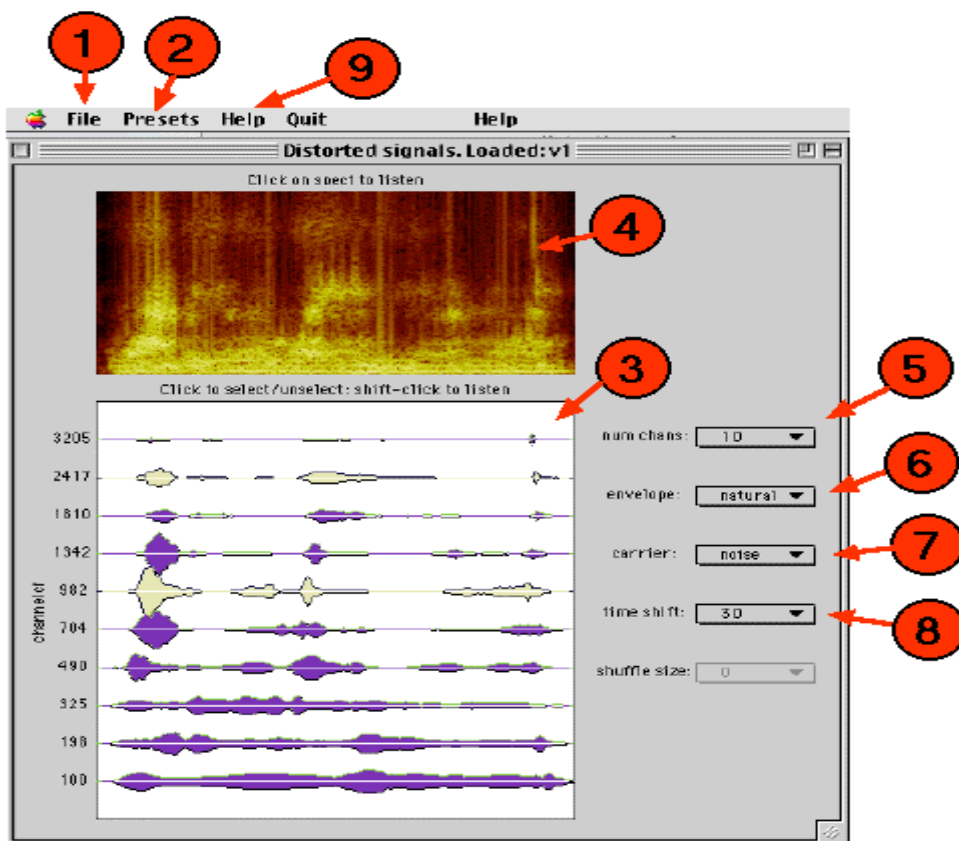
<< Η ομιλία >> συχνά φτάνει στα αυτιά του ακροατή σαν μια μάλλον διαφορετική υπόσταση από ότι ήταν κατά την παραγωγή της. Τα αποτελέσματα της αντίληψης και η παρουσία άλλων πηγών συνεισφέρουν στο αλλοιωμένο σήμα . Οι ακροατές κατέχουν στρατηγικές για να χειριστούν πολλών ειδών παραμορφώσεις και η αποκάλυψη αυτών των στρατηγικών είναι σημαντική για τον σχεδιασμό εύρωστων συστημάτων για computation hearing.

Το πεδίο των δυνατών παραμορφώσεων είναι αρκετά ευρύ και πολλών ειδών τροποποιήσεις σήματος έχουν αποτελέσει το αντικείμενο ψυχοακουστικής έρευνας από το έργο του Fletcher στην αρχή του εικοστού αιώνα (δες Fletcher 1953 ; Allen , 1994). Τα σήματα ομιλίας έχουν υποβληθεί σε τέτοιες παραμορφώσεις όπως :

- Φασματικό φιλτράρισμα .
- Φασματική εξασθένιση σε μικρό αριθμό ημιτονοειδών (δες την sine-wave speech demo) ή τις περιορισμένες ζώνες συχνοτήτων.
- Πολυκαναλική πρόσκαιρη παραμόρφωση .
- Περιβάλλουσες τροποποιήσεις και τροποποιήσεις μεταφοράς .
- Περιοδικές ή τυχαίες διακοπές και σβήσιμο (δες πρόσκαιρη επαγωγή demo) .

Η παρούσα παρουσίαση επιτρέπει στον χρήστη να παράγει μερικές παραμορφώσεις αυτού του τύπου .

Η Παρουσίαση



Για να ξεκινήσει η παρουσίαση πληκτρολογήστε << παραμόρφωση >> στο Matlab Prompt . Θα εμφανιστεί ένα παράθυρο ως ανωτέρω . Επιλέξτε την επιλογή Load στο File Menu (1) για να διαλέξετε ένα αρχείο ήχου. (Μερικά παραδείγματα ομιλίας και μουσικής παρέχονται με την κατανομή , αλλά μπορούν να φορτωθούν οποιαδήποτε αρχεία με επέκταση .au ή sdh . Ένα φασματόγραμμα του ήχου κάνει την εμφάνιση του στο πάνω μέρος του panel (4). Καθώς εφαρμόζονται παραμορφώσεις , αυτή η διάταξη αναβαθμίζεται για να δείξει το φασματόγραμμα του παραμορφωμένου ήχου. Για να ακούσετε το σήμα κάντε κλικ οπουδήποτε στο φασματόγραμμα . Σύντομα μετά την εμφάνιση του φασματογράμματος , θα παρουσιαστεί μια σειρά από κυματομορφές στο κάτω μέρος του panel (3). Αυτές αντιστοιχούν στις περιβάλλουσες Hilbert της απόδοσης μιας τράπεζας ακουστικών gammatone φίλτρων . Οι κεντρικές συχνότητες των φίλτρων δείχνονται στην αριστερή πλευρά της διάταξης. Συνίσταται να περιορίσετε τον αριθμό των επιλεγμένων καναλιών σε 10 ή και σε λιγότερα εκτός και αν ασχολείστε με σύντομα σήματα (< 1 sec) . Σε μερικές πλατφόρμες η μνήμη και η ταχύτητα υπολογισμού έχει σημασία όταν έχουμε να κάνουμε με μεγαλύτερο αριθμό καναλιών .

ΠΑΡΑΤΗΡΗΣΗ : Είναι σημαντικό να παρατηρήσετε πως το εύρος των συχνοτήτων του φίλτρου δεν ρυθμίζεται για να εξασφαλίσει ίση κάλυψη όλων των φασματικών περιοχών . Τα πλάτη των συχνοτήτων κινούνται προς εκείνα που όρισε ο Glasberg και ο Moore (1990) και αντιπροσωπεύουν εκτιμήσεις της αποτελεσματικής , εξαρτημένης από την συχνότητα ανάλυσης , της ακουστικής περιφέρειας . Μελλοντικές εκδοχές ίσως επιτρέψουν κατάλληλη επεξεργασία αυτού του θέματος , αλλά η πλεονασμός του σήματος ομιλίας εξασφαλίζει πως ακόμη και μια ανάλυση 8 gammatone δίνει μια τελείως κατάλληλη καθαρή βάση για επακόλουθες παραμορφώσεις.

Η διάταξη του φασματογράμματος έχει ένα γραμμικό in-Hz Y-άξονα , ενώ οι κεντρικές συχνότητες του φίλτρου (C F) συντάσσονται σε μια ERB-rate κλίμακα. Η τελευταία είναι κατά προσέγγιση λογαριθμική.

Μπορείτε να ακούσετε μεμονωμένες ζώνες συχνοτήτων ολισθαίνοντας με το ποντίκι και κάνοντας κλικ πάνω στην κυματομορφή τους .(Σε πλατφόρμες άλλες από την Mac ίσως αυτό συνεπάγεται χρήση του δεξιού πλήκτρου του ποντικιού). Κάνοντας κλικ σε ατομικές κυματομορφές επιλέγει / αποεπιλέγει εκείνο το σήμα. Ένα σήμα που δεν είναι επιλεγμένο δεν συνεισφέρει στο τελικό αποτέλεσμα.

Επιλέγοντας ή αποεπιλέγοντας μπορείτε να εξερευνήσετε ποικίλες μορφές φασματικού φιλτραρίσματος. Αντιστρόφως χρησιμοποιήστε το presets menu (2) για να επιταχύνετε την διαδικασία επιλογής.

Άλλες παραμορφώσεις ρυθμίζονται από τα popup menus στα δεξιά της διάταξης. Το popup menu (6) επιτρέπει στην περιβάλλουσα σε κάθε ζώνη συχνοτήτων να αντικατασταθεί από μια κυματομορφή θορύβου , ή από μία σταθερή (ρυθμίστε στο μέσον της περιβάλλουσας). Το popup menu (7) επιτρέπει στον φορέα να είναι ένα σήμα θορύβου ή ένας τόνος. Το σήμα θορύβου προκύπτει περνώντας ένα θόρυβο ευρείας ζώνης συχνοτήτων διαμέσου του κάθε gammatone φίλτρου. Η συχνότητα του τόνου ρυθμίζεται στο κανάλι CF.

Το popup menu (8) προσδιορίζει έναν μέγιστο χρόνο ολίσθησης (σε ms) που πρέπει να εφαρμοστεί σε κάθε κανάλι . Η τελική χρονική ολίσθηση που εφαρμόστηκε είναι μια ψευδοτυχαία καθυστέρηση περιορισμένη από αυτόν τον αριθμό. Σ' όλες τις περιπτώσεις και οι δύο διατάξεις αναβαθμίζονται για να δείξουν το αποτέλεσμα της παραμόρφωσης. Οι παραμορφώσεις μπορεί να συνδυαστούν με αυθαιρέτους τρόπους. Περιβάλλουσα και τρέχουσα μεταβολή σε κάθε ζώνη

συχνότητας εφαρμόζονται ανεξάρτητα και η προκύπτουσα κυματομορφή σε κάθε ζώνη είναι προσωρινά παραμορφωμένη .
Υπάρχει άφθονος χώρος να προσθέσετε επιπλέον παραμορφώσεις .

Θέματα προς έρευνα

Αφού είναι εύκολο να ακούσετε μέχρι τέλους εκφράσεις λόγου άπαξ και γνωρίζετε τι είναι , συνίσταται να φορτώνετε διαφορετικές εκφράσεις συχνά και να ξεκινάτε με τις πιο προκλητικές συνθήκες , σταδιακά εισάγοντας περισσότερες πληροφορίες , μέχρι που η έκφραση λόγου να μπορεί αμέσως να αναγνωριστεί.

1. Βασικό φασματικό φιλτράρισμα. Ακούστε το αποτέλεσμα του βαθυπερατού φιλτραρίσματος , του υψιπερατού φιλτραρίσματος , του ζωνοπερατού φιλτραρίσματος και του bandstop filtering. Η τελευταία συνθήκη έχει ερευνηθεί πρόσφατα από τον Lippmann (1966) για την αναγνώριση των συμφώνων.
2. Χωριστές ζώνες συχνότητων . Ακούστε πρώτα την χαμηλότερη ζώνη συχνότητας, ύστερα την ύψιστη , κατόπιν κάποια στις μεσαίες συχνότητες. Οι Warren et al (1995) πρόσφατα μέτρησαν την απόδοση αναγνώρισης λέξεων για άκρως περιορισμένες ζώνες συχνότητων σαν λειτουργία του CF τους και βρήκαν πολύ καλύτερη απόδοση στα 1500 Hz απ' ό τι στα 300 ή 6000 Hz. Εν τούτοις ...
3. Δύο ζώνες συχνότητωνΟι Warren et al βρήκαν υπερπροσθετική απόδοση όταν οι χαμηλές και υψηλές ζώνες συχνότητας παρουσιάζονται μαζί .
4. Παραμόρφωση φορέως. Ακούστε το αποτέλεσμα αντικατάστασης φορέων από θόρυβο και από τόνους . Εξετάστε το φασματογράφημα για να δείτε πως διασπώνται οι πληροφορίες που σχετίζονται με την δομή φώνησης και φωνοσυντονισμού. Η κατάσταση φορέας θορύβου σχετίζεται με μια πρόσφατη ανακάλυψη από τον Shannon et al (1995) , αν και αυτή συμπεριλάμβανε ένα μικρότερο αριθμό ευρύτερων ζωνών συχνότητων.
5. Παραμόρφωση περιβάλλουσας.
6. Πρόσκαιρη παραμόρφωση. Αρχίζοντας με τις πιο μακροχρόνιες καθυστερήσεις (1 sec) , σταδιακά μειώστε το ποσόν της πρόσκαιρης παραμόρφωσης μέχρι που να μπορεί να αναγνωριστεί η έκφραση λόγου . Αυτή η τροποποίηση σήματος σχετίζεται με το πρόσκαιρο φιλτράρισμα διαμόρφωσης του Drullman (1995).
7. Είναι διδακτικό να συγκρίνετε τα αποτελέσματα ταυτόσημων παραμορφώσεων σε σήματα άλλα εκτός της ομιλίας. Μερικά μουσικά παραδείγματα παρέχονται στην κανονική συνήθη διανομή.

Αναφορές

1. Allen (1994) . *IEEE Trans. Speech & Audio Proc.* , 2(4), 567-577.
2. Cooke (1996) . *Proc. ESCA Workshop on Auditory Basis for Speech Perception*, Keele.
3. Drullman (1995). *JASA*, **97**(1), 585-592.
4. Fletcher (1953). *Speech & Hearing in Communication*. Van Nostrand.
5. Glasberg & Moore (1990). *Hearing Research*, **47**, 103-138.
6. Lippmann (1996). *IEEE Trans. on Speech & Audio Processing*, **4**(1), 66-69.
7. Shannon et al (1995). *Science*, **270**, 303-304.
8. Warren et al (1995). *Perc. & Psychophys.*, **57**(2), 175-182

Περαιτέρω μελέτη

- Η δυνατότητα του ακουστικού συστήματος να χειριστεί φασματικό φιλτράρισμα και εξασθένιση είναι μία από τις πηγές έμπνευσης για απώλεια δεδομένων στην αναγνώριση της ομιλίας. Δες το demo για περισσότερες αναφορές σε αυτό το σύγγραμμα.
- Η αντοχή των ακροατών σε μια πολυκαναλική χρονική παραμόρφωση είναι ένας από τους παράγοντες που ενεργούν σαν κίνητρο πίσω από την σύνθετη ροή που προσεγγίζει το δυνάμωμα του ASR. Δες Bourlard & Dupont (1977). ICASSP ' 97, 1251-1254.