

ΕΛΛΕΙΠΟΝΤΑ ΔΕΔΟΜΕΝΑ ASR

ΕΙΣΑΓΩΓΗ

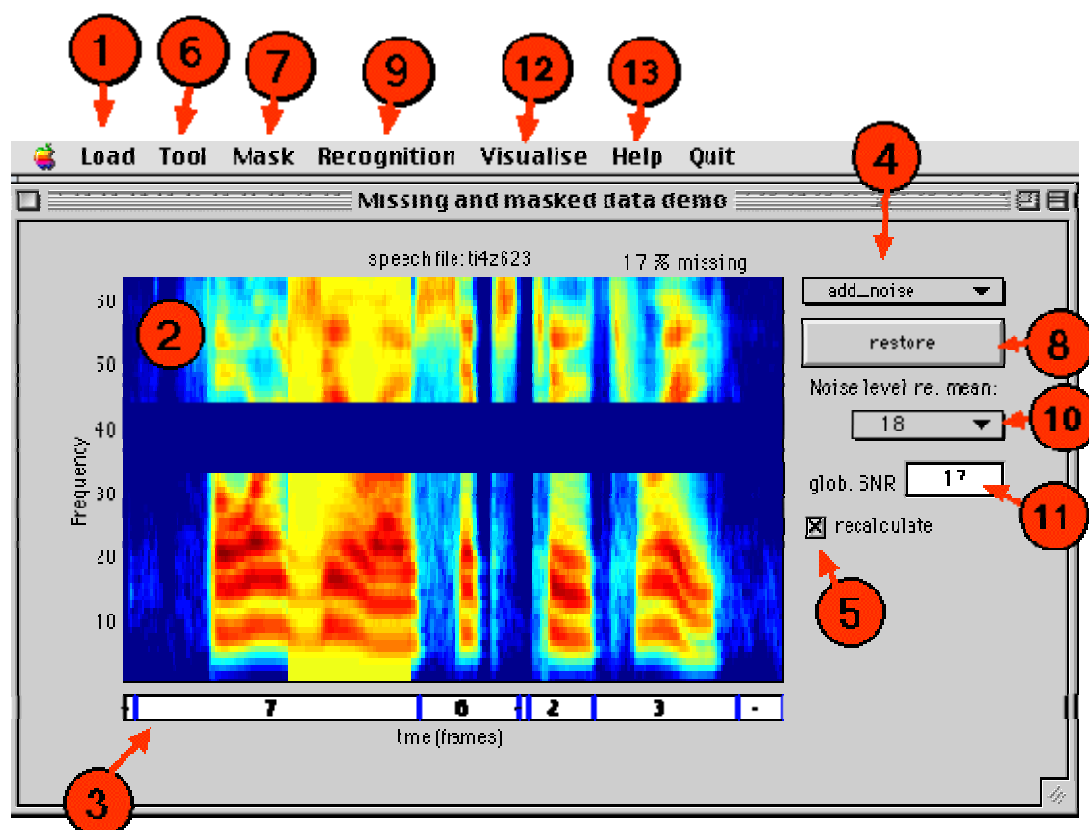
Οι ακροατές είναι ικανοί να αντιλαμβάνονται την ομιλία παρά τις αρκετά σοβαρές παραμορφώσεις (δείτε την παρουσίαση παραμορφωμένη ομιλία). Οι κοινές όψεις πολλών παραμορφώσεων περιλαμβάνουν την απουσία φασματο-χρονικών περιοχών ή την παρουσία πρόσθετου θορύβου. Τα ελλείποντα δεδομένα εμφανίζονται φυσικά σε συνθήκες τέτοιες όπως η τηλεφωνική συνδιάλεξη και η πτώση έντασης καναλιού. Η αναγνώριση ομιλίας ελλειπόντων δεδομένων είναι μια πρόσφατη προσέγγιση (Cooke et al, 1994) στο εύρωστο ASR βασιζόμενη στη λύση των ακόλουθων υποπροβλημάτων :

- Αναγνώριση αξιόπιστων ενδείξεων και
- Προσαρμογή αναγνωριστών για να χειριστούν τα ελλείποντα και αποκρυμμένα δεδομένα.

Οι λύσεις στο πρώτο υποπρόβλημα κυμαίνονται από τεχνικές μείωσης υπάρχοντος θορύβου τέτοιες όπως φασματική αφαίρεση και τοπική SNR εκτίμηση (Hirsch 1993) μέχρι πιο γενικού στόχου προσεγγίσεις τέτοιες όπως είναι η ανάλυση ακουστικής σκηνής (Rosenthal and Okuno, 1998).

Μια προσέγγιση στο δεύτερο πρόβλημα επιδεικνύεται εδώ. Ο σκοπός της παρουσίασης είναι να επιτρέψει στον χρήστη να εξερευνήσει το αποτέλεσμα στην αναγνώριση της μετακίνησης των φασματο-χρονικών περιοχών ή της πρόσθεσης θορύβων στην ομιλία. Αξίζει να σημειωθεί ότι το υποκείμενο HMM σύστημα είναι πολύ απλό (απλή μίξη συστατικών) και κάνει λάθη ακόμη και χωρίς τα ελλείποντα δεδομένα.

Η Παρουσίαση



Πληκτρολογήστε «md» στο MATLAB prompt. Χρησιμοποιείστε το υπομενού ομιλίας του μενού φόρτωσης (1) για να επιλέξετε ένα αρχείο και να το φορτώσετε. Τα αρχεία ομιλίας που παρέχει η παρουσίαση είναι χαρακτήρες από ψηφία (μήκος 4 ή 5) από το TIDigits corpus (ref). Η ταυτότητα της ακολουθίας των ψηφίων είναι εμφανής από το όνομα της.

[Παρατήρηση : Τα ονόματα των αρχείων είναι της μορφής ti 9082, που υποδηλώνει ότι αυτό είναι το σύνολο «εννέα-μηδέν-οκτώ-δύο», από το TIDigits corpus. Μοντέλα αναγνώρισης υπάρχουν για τα 9 ψηφία 1-9, μαζί με το « ο » για το «oh» , « z » για το «μηδέν» και το «--» που δείχνει σιωπή.

Για αυτήν την παρουσίαση η αναγνώριση επιτελείται χρησιμοποιώντας μάλλον απλή μεμονωμένη μίξη HMMs χωρίς χρήση παραγώγων, ούτως ώστε η εκτέλεση δεν είναι τέλεια ακόμη και χωρίς διαγραφές !]

Κατά την φόρτωση , το σύστημα παρουσιάζει ένα φασματογράφημα δραστηριοποιούμενο ακουστικά, που χρησιμοποιείται σαν μοναδική βάση για την αναγνώριση (2). Μετά από λίγο (περισσότερο για τα αργά συστήματα, αλλά λίγα δευτερόλεπτα για ένα PowerMac που έχει ταχύτητα 200 MHz) το αναγνωρισμένο αποτέλεσμα θα φανεί στο κάτω πάνελ (3). Τα σύνορα που εκτίθενται είναι εκείνα που επιτυγχάνονται ανιχνεύοντας προς τα πίσω διαμέσου της καλύτερης ακολουθίας μοντέλων κατά τον συνήθη τρόπο.

Διόρθωση της έκθεσης

Σ' αυτό το σημείο, κάνοντας κλικ οπουδήποτε στην φασματογραφική εικόνα, έχει σαν αποτέλεσμα την επέμβαση αλλαγή και διόρθωση του ακουστικού κειμένου που καθορίζεται από το popup μενού (4). Για παράδειγμα, εάν το mode είναι διαγραφή, μια περιοχή κάτω από τον cursor μετακινείται από το display κάθε φορά που ο χρήστης κάνει κλικ στο φασματογράφημα. Σ' αυτό το σημείο αξίζει να παρατηρήσουμε ότι η παρουσίαση δρα σαν λογιστικό φύλλο.

Οι αλλαγές κάνουν να ξαναρχίσει ο αναγνωριστής που αναβαθμίζει το αναγνωρισμένο αποτέλεσμα «σχεδόν στιγμιαία». Αν χρησιμοποιείτε αργό μηχανήμα (που εδώ ορίζεται σαν κάποιο που παίρνει περισσότερο από λίγα δευτερόλεπτα για να εκτελέσει την αναγνώριση), ίσως επιθυμείτε να κάνετε untick στο recalculate checkbox (5) μέχρι που να θελήσετε να γίνει η αναγνώριση.

Η έκταση της περιοχής που επηρεάζεται από το κλικ που κάνετε στο display ρυθμίζεται από το tool μενού (6). Εδώ μπορείτε να επιλέξετε την περιοχή που επηρεάζεται από έναν χειρισμό επέμβασης αλλαγής προσδιορίζοντας την περιοχή συχνότητας (στα κανάλια) και την διάρκεια (σε χρονικά πλαίσια 10 ms).

Άλλες επιλογές επέμβασης αλλαγής είναι το add noise (πρόσθεση θορύβου) που προσθέτει θόρυβο στο προσδιορισμένο επίπεδο (σχετικό με το μέσο σήμα) με το popup μενού (10) και το restore το οποίο αποκαθιστά τις αρχικές αξίες. Επιπλέον όλο το display μπορεί να ξανατεθεί οποιαδήποτε στιγμή πατώντας το κουμπί restore (8). Η μετατόπιση ή αποκατάσταση των δεδομένων έχει εμφανές αποτέλεσμα στο display. Μια status line πάνω δεξιά του display δείχνει το ποσοστό των ελλειπόντων χρονοσυχνοτικών περιοχών.

Ένα οπτικό αποτέλεσμα της πρόσθεσης θορύβου εξαρτάται από το επίπεδο θορύβου που είναι σχετικό με το υπόβαθρο της ομιλίας. Το σφαιρικό SNR αλλάζει από οποιαδήποτε προσθήκη θορύβου και υποδεικνύεται στο display. Επιπροσθέτως ο χρήστης μπορεί να θέσει ένα ζητούμενο SNR και το υπόβαθρο του θορύβου θα τροποποιηθεί για να ικανοποιήσει αυτό το κριτήριο (11).

Αναγνώριση

Ο τρόπος κατά τον οποίο τα ελλείποντα και αποκρυμμένα δεδομένα αναγνωρίζονται, ελέγχεται από τις επιλογές που είναι διαθέσιμες στα (7) mask και (9) recognition μενού. Ο προεπιλεγμένος αναγνωριστής είναι μια συμβατική στρατηγική η οποία μεταχειρίζεται τα δεδομένα σαν να είναι πλήρη. Δηλαδή περιοχές που έχουν διαγραφεί από τον χρήστη χειρίζονται ως έχουσες μηδενική αξία (ενέργεια). Ομοίως ο πρόσθετος θόρυβος αποτελεί μέρος των δεδομένων που δίνονται στον αναγνωριστή. Θέτοντας την επιλογή ελλείποντα δεδομένα μέσω του μενού αναγνώρισης οδηγεί τον αναγνωριστή να μεταχειριστεί ορισμένες περιοχές σαν να απουσιάζουν. Στην απλούστατη περίπτωση, αυτές οι περιοχές είναι απλώς εκείνες τις οποίες έχει διαγράψει ο χρήστης από το display. Εν τούτοις, για την περίπτωση του πρόσθετου θορύβου, ο χρήστης μπορεί να αναγκάσει τον αναγνωριστή να μεταχειριστεί εκείνες τις περιοχές μ' ένα αρνητικό τοπικό SNR σαν ελλείπουσες, επιλέγοντας την $snr > 0$, επιλογή μέσω του mask μενού. Το display αμέσως αναβαθμίζεται για να αποκαλύψει εκείνες τις περιοχές που ικανοποιούν αυτό το κριτήριο. Μια περαιτέρω επιλογή αναγνώρισης είναι να χρησιμοποιήσετε απλά φασματικές κορυφές (ή το υποσύνολο τους που δεν απουσιάζει ούτε συγκαλύπτεται από τον θόρυβο) στην διαδικασία της αναγνώρισης. Αυτή η επιλογή επιτυγχάνεται μέσω της επιλογής peaks στο μενού masks. Και πάλι το display αναβαθμίζεται για να δείξει τα σημεία επί των οποίων συμβαίνει η αναγνώριση.

Επιθεωρώντας τις HMMS και OUTPUT πιθανότητες

Ορισμένες όψεις των μοντέλων αναγνώρισης και διαδικασίας μπορούν να απεικονιστούν (12) :

- Απεικόνιση > HMMS σχεδιάζει τα μέσα / διακυμάνσεις σε κάθε HMM κατάσταση.
- Απεικόνιση > output probs εκθέτει τις πιθανότητες εξόδου για κάθε πλαίσιο. Αυτό ενημερώνεται όταν γίνονται αλλαγές στην είσοδο του αναγνωριστή.

Προ-ρύθμιση διαγραφών και θόρυβος

Είναι δυνατόν να εφαρμοστούν ορισμένα μοντέλα διαγραφής χωρίς την χρήση ποντικιού για να επιλέξετε περιοχές. Αυτά είναι διαθέσιμα στο υπομενού load deletions και περιλαμβάνουν :

- Τυχαία διαγραφή στο 50% και 90% επίπεδο.
- Τυχαία μετακίνηση πλαισίων ή καναλιών στο 50% ή στο 90%.
- Μετακίνηση όλων των σημείων χαμηλής ενέργειας.
- Βαθυπερατό, ζωνοπερατό ή υψιπερατό φιλτράρισμα.
- Μετακίνηση όλων εκτός από ένα ζεύγος μεμονωμένης περιοχής συχνοτήτων.

Επιπροσθέτως, τα μοντέλα θορύβου (προς το παρόν, αυτά είναι απλά άλλα αρχεία ομιλίας) μπορεί να φορτωθούν από το load > noise υπομενού. Στο εγγύς μέλλον, η παρουσίαση θα εμπλουτιστεί ώστε να περιλαμβάνει μια σειρά από τεχνικές automatic local SNR.

Θέματα προς έρευνα

1. Φορτώστε ένα αρχείο ομιλίας και παρατηρήστε το αποτέλεσμα αναγνώρισης. Εφαρμόστε ένα μοντέλο προρυθμισμένης διαγραφής. Ξανά, κοιτάξτε τι συμβαίνει στην λειτουργία της αναγνώρισης. Ανοίξτε το missing data recognition . Αυτό θα έπρεπε να βελτιώσει την λειτουργία του αναγνωριστή.
2. Εξετάστε το αποτέλεσμα της πρόσθεσης θορύβου. Ξανά, φορτώστε ένα αρχείο ομιλίας, αλλά αυτή την φορά προσθέστε μερικά «μπαλώματα» θορύβου με το ποντίκι. Η αναγνώριση θα πρέπει να υποβαθμιστεί (αν όχι, δεν προσθέτετε αρκετό θόρυβο !). Κατόπιν, θέστε το display στο $snr > 0$ από το mask μενού και ανοίξτε το missing data recognition. Αυτό εξασφαλίζει ότι το σύστημα χρησιμοποιεί μόνο εκείνα τα σημεία των οποίων το SNR είναι θετικό.

