



ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

**ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΑ
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«ΑΥΤΟΜΑΤΟΠΟΙΗΜΕΝΗ ΗΧΗΤΙΚΗ ΕΠΕΝΔΥΣΗ ΡΟΩΝ
ΣΤΑΤΙΚΩΝ ΕΙΚΟΝΩΝ»**

ΚΑΤΑΧΑΝΑΣ ΜΙΧΑΗΛ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ

ΦΛΩΡΟΣ ΑΝΔΡΕΑΣ

ΠΑΤΡΑ

ΙΟΥΛΙΟΣ, 2013

© ΕΑΠ., 2013

Η παρούσα διατριβή η οποία εκπονήθηκε στα πλαίσια της ΘΕ “Διπλωματική Εργασία” του προγράμματος “Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα” (ΠΛΗΣ), και τα λοιπά αποτελέσματα της αντίστοιχης Διπλωματικής Εργασίας (ΔΕ) αποτελούν συνιδιοκτησία του ΕΑΠ και του φοιτητή, ο καθένας από τους οποίους έχει το δικαίωμα ανεξάρτητης χρήσης και αναπαραγωγής τους (στο σύνολο ή τμηματικά) για διδακτικούς και ερευνητικούς σκοπούς, σε κάθε περίπτωση αναφέροντας τον τίτλο και το συγγραφέας και το ΕΑΠ, όπου εκπονήθηκε η Διπλωματική Εργασία, καθώς και τον επιβλέποντα και την επιτροπή κρίσης.

“ΑΥΤΟΜΑΤΟΠΟΙΗΜΕΝΗ ΗΧΗΤΙΚΗ ΕΠΕΝΔΥΣΗ ΡΟΩΝ ΣΤΑΤΙΚΩΝ ΕΙΚΟΝΩΝ”

Καταχανάς Μιχαήλ

Φλώρος Ανδρέας

Επιβλέπων

Ιωάννου Κωνσταντίνος

Μέλος Επιτροπής 1^ο

ΠΕΡΙΛΗΨΗ

Ο συσχετισμός εικόνας και ήχου στην καθημερινή ζωή αποτελεί μian αυτονόητη διαδικασία για τα υγιή ενήλικα άτομα. Πρόκειται ωστόσο για μια λειτουργία η οποία περιλαμβάνει τη συμμετοχή των πλέον εξελιγμένων αισθήσεων του ανθρώπου και που προϋποθέτει τη σωστή λειτουργία μεγάλου αριθμού μηχανισμών τόσο σε φυσικό και βιοχημικό επίπεδο. Αποτελεί, δε, πεδίο μακρόχρονης και εντατικής έρευνας και επιστημονικού πειραματισμού και βρίθει εφαρμογών στην ιατρική, την εκπαίδευση, την τέχνη και την ψυχαγωγία. Η παρούσα διπλωματική εργασία επικεντρώνεται στο σχεδιασμό και την υλοποίηση λογισμικού που εστιάζει σε μια από τις εκφάνσεις του συνδυασμού των δυο αυτών αισθήσεων. Συγκεκριμένα, ασχολείται με την ανάπτυξη εφαρμογής η οποία παράγει αυτόματα, σε πραγματικό χρόνο, ηχητικές επενδύσεις για ροές στατικών εικόνων χρησιμοποιώντας δεδομένα που εξάγει από τις προβαλλόμενες εικόνες.

Η εργασία περιλαμβάνει, μεταξύ άλλων, μian εισαγωγή στον τρόπο με τον οποίο ο άνθρωπος και γενικότερα τα ανώτερα θηλαστικά αντιλαμβάνονται και συσχετίζουν τα ερεθίσματα που προσλαμβάνονται μέσω της οπτικής οδού με εκείνα που γίνονται αντιληπτά μέσω της ακοής. Επίσης αναφέρεται στις διάφορες εφαρμογές που βρίσκει το συγκεκριμένο γνωστικό πεδίο σε ποικίλους τομείς. Στη συνέχεια παρατίθεται ανάλυση για την εφαρμογή του στο χώρο της τέχνης και γενικότερα της ψυχαγωγίας και διασκέδασης.

Σχετικά με το σύστημα λογισμικού που σχεδιάστηκε και αναπτύχθηκε στα

πλαίσια της εργασίας, παρουσιάζονται οι αρχές σχεδίασης και ανάπτυξης του καθώς επίσης και τα εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξη και υλοποίησή του, λ.χ. η γλώσσα προγραμματισμού Vala και το λογισμικό Open Sound Control (OSC). Επίσης αναλύονται οι αλγόριθμοι που χρησιμοποιήθηκαν και η αξιολόγησή τους από πλευράς αποδοτικότητας.

Αναφέρεται επίσης ο τρόπος με τον οποίο χρησιμοποιήθηκαν οι υπάρχουσες εφαρμογές ανοικτού κώδικα Shotwell (για τη διαχείριση και απεικόνιση βάσεων δεδομένων με φωτογραφίες) και Csound (λογισμικό παραγωγής και επεξεργασίας ήχων) στα πλαίσια της υλοποίησης. Επίσης αναλύονται οι παράγοντες οι οποίοι συντέλεσαν στις διάφορες αποφάσεις σχεδιαστικής προσέγγισης που πάρθηκαν κατά τη διάρκεια ανάπτυξης. Επίσης γίνεται σύντομη αναφορά στο εργαλείο ανάλυσης και σύλληψης (capturing) και ανάλυσης πακέτων δικτύων υπολογιστών Wireshark όπως επίσης και οι προκλήσεις που προέκυψαν κατά την ανάπτυξη. Αναφέρονται τα συμπεράσματα από την ανάπτυξη και χρήση του λογισμικού και δίνονται μερικές διαπιστώσεις για την χρησιμοποίηση των δεδομένων που εξάγονται από μια φωτογραφία για την παραγωγή ήχου. Τέλος προτείνονται κάποιες εκδοχές περαιτέρω ανάπτυξης της υπάρχουσας εφαρμογής ή παρόμοιων εφαρμογών λογισμικού.

Λέξεις-κλειδιά: Ροή στατικών εικόνων, Ηχητική επένδυση, Εφαρμογή πραγματικού χρόνου, Vala, OSC, Shotwell, Csound, QuteCsound, Linux, Open Source, Wireshark.

ABSTRACT

The correlation of image and sound in everyday life is considered to be an obvious procedure for healthy adults. However, it is an operation which involves the participation of the most sophisticated human senses and requires the efficient cooperation of a large number of mechanisms in both natural and biochemical level. Moreover, it consists a field of long-running research and scientific experimentation and has resulted in numerous applications in medicine, education, art and entertainment. This thesis focuses on the design and implementation of software that combines these two senses. More specifically, it describes the design and development process of a software application which automatically generates audio, in real time, based on input extracted

from static images during projecting these images in a slideshow.

This work includes an introduction to the way humans and primates perceive stimuli obtained via optical pathway and those received through hearing. It also describes the various applications of the related principles in various scientific and experimentalization fields. An analysis of the application in the field of art and entertainment in general is also provided.

When it comes to the software application which was designed and implemented for this thesis, the concepts and guidelines along with the utilized tools, such as programming language Vala and Open Sound Control (OSC) which have been used during the process are described. The algorithms used for implementing the sound generation part are examined from both conceptual and execution times point of view and also a brief description is provided regarding the way they were used for the generation of the sound.

Next, the way this application took advantage of existing open source applications Shotwell (catering for the image processing and sound data extraction) and Csound (a programming language dedicated for sound processing and sound production) is explained. A reference is also made also to the several different aspects and challenges which affected the decisions taken during the design of the application. A short reference is also made to the network protocols capturing and analysis software Wireshark and how this was used in the scope of this thesis.

Finally, some conclusions drawn from the project are listed and proposals for future development are discussed.

Keywords: slideshow, static image, soundtrack, realtime, Vala, OSC, Linux, Open Source, Shotwell, Csound, QuteCsound, Linux, Wireshark.

1	ΕΙΣΑΓΩΓΗ	8
	ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΣΕ ΕΦΑΡΜΟΓΕΣ ΑΥΤΟΜΑΤΟΠΟΙΗΜΕΝΗΣ ΜΕΤΑΤΡΟΠΗΣ ΔΕΔΟΜΕΝΩΝ ΕΙΚΟΝΑΣ ΣΕ ΗΧΟ	8
	ΣΚΟΠΟΣ ΚΑΙ ΔΟΜΗ	15
2	ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	16
	ΟΙ ΑΙΣΘΗΣΕΙΣ ΤΗΣ ΟΡΑΣΗΣ ΚΑΙ ΤΗΣ ΑΚΟΗΣ	16
	ΣΚΟΠΟΙ ΚΑΙ ΠΡΟΣΔΟΚΩΜΕΝΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	31
3	ΑΝΑΛΥΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ	33
	ΕΙΣΑΓΩΓΗ	33
4	ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΑΠΤΥΞΗΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ	42
	ΕΙΣΑΓΩΓΗ	42
5	ΣΧΕΔΙΑΣΗ ΣΥΣΤΗΜΑΤΟΣ	48
	ΠΛΑΤΦΟΡΜΑ ΥΛΟΠΟΙΗΣΗΣ	50
	ΕΡΓΑΛΕΙΑ ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΚΤΕΛΕΣΗΣ ΚΑΙ ΕΠΑΝΑΧΡΗΣΙΜΟΠΟΙΗΣΗ ΥΠΑΡΧΟΥΣΩΝ ΕΦΑΡΜΟΓΩΝ	53
	ΑΡΧΙΤΕΚΤΟΝΙΚΗ	54
6	ΑΠΟΤΕΛΕΣΜΑΤΑ	65
	ΔΙΑΔΙΚΑΣΙΑ ΔΟΚΙΜΩΝ	65
	ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΑΚΤΙΚΗΣ ΕΦΑΡΜΟΓΗΣ ΕΡΓΑΛΕΙΩΝ	66
	ΑΠΟΤΕΛΕΣΜΑΤΑ ΧΡΗΣΗΣ	71

7	ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ	72
	ΣΥΜΠΕΡΑΣΜΑΤΑ	73
	ΠΡΟΟΠΤΙΚΕΣ ΑΝΑΠΤΥΞΗΣ	74
8	ΒΙΒΛΙΟΓΡΑΦΙΑ	76
9	ΠΑΡΑΡΤΗΜΑ Α	82
10	ΠΑΡΑΡΤΗΜΑ Β	84

1 ΕΙΣΑΓΩΓΗ

Η εξέλιξη των ηλεκτρονικών υπολογιστών και η δυνατότητα διαχείρισης ψηφιακής πληροφορίας έδωσε στους χρήστες και τους προγραμματιστές τη δυνατότητα να αξιοποιήσουν το τεχνολογικό μέσο σε χώρους έξω από τα αυστηρά πλαίσια της εργασίας. Έδωσε, δε, το έναυσμα για τη μελέτη και ανάπτυξη εφαρμογών με στόχο την οπτική και ακουστική ικανοποίηση του χρήστη αλλά και τη δημιουργία συλλογών από αρχεία σχετιζόμενων με τη διασκέδαση και ψυχαγωγία του χρήστη όπως για παράδειγμα ψηφιακών φωτογραφιών και αρχείων μουσικής. Αργότερα στάθηκε η αφορμή και για πειραματισμό πάνω στους διαφορετικούς τρόπους διαχείρισης των διαφόρων αρχείων.

Ιστορική Αναδρομή σε Εφαρμογές Αυτοματοποιημένης Μετατροπής Δεδομένων Εικόνας σε Ήχο

Η εξαγωγή στοιχείων από μian εικόνα και η περαιτέρω χρήση τους για την παραγωγή ήχων ή/και αρμονικών φθόγγων (και υπό μian έννοια μουσικής) δεν είναι κάτι καινούργιο. Μια μη-αυτοματοποιημένη έκδοση αυτής της σύλληψης φυσικά αποτελούσε και αποτελεί ο κινηματογράφος και οι τεχνικές ηχητικής επένδυσης της. Με πρόγονο της την ηχητική επένδυση των φιλμ βωβού κινηματογράφου. Κατά τη διάρκεια προβολής τους, σχεδόν πάντοτε μουσικοί φρόντιζαν για την ζωντανή ηχητική/μουσική επένδυσή της [7].



Εικόνα 1 Οι πρώτες προσπάθειες επένδυσης εικόνας με ήχο

Αργότερα εφευρέθηκαν και χρησιμοποιήθηκαν μέθοδοι και για την παραγωγή “φυσικών” ήχων σχετικών με την πλοκή του έργου για παράδειγμα για την απόδοση του καλπασμού αλόγου ή της βροντής. Η αυτοματοποιημένη παραγωγή ηχητικής επένδυσης για εικόνες και γενικότερα η εξαγωγή δεδομένων ήχων από εικόνα δύο ή τριών διαστάσεων είναι σαφώς πιο πρόσφατη ως σύλληψη. Φυσικά στη σύλληψη και υλοποίηση μιας τέτοιας φαινομενικά ανορθόδοξης ιδέας συντέλεσαν καταλυτικά οι ηλεκτρονικοί υπολογιστές. Κάτι τέτοιο είναι αναμενόμενο, μιας και η δυνατότητα ψηφιακής κωδικοποίησης αποτελεί αναγκαία συνθήκη για τη δυνατότητα εξαγωγής δεδομένων που θα μπορούσαν να αξιοποιηθούν για την παραγωγή μη οπτικού ερεθίσματος από αυτή αλλά και το αντίστροφο (ήχος αποθηκευμένος σε ψηφιακή μορφή χρησιμοποιείται για την παραγωγή οπτικού ερεθίσματος).

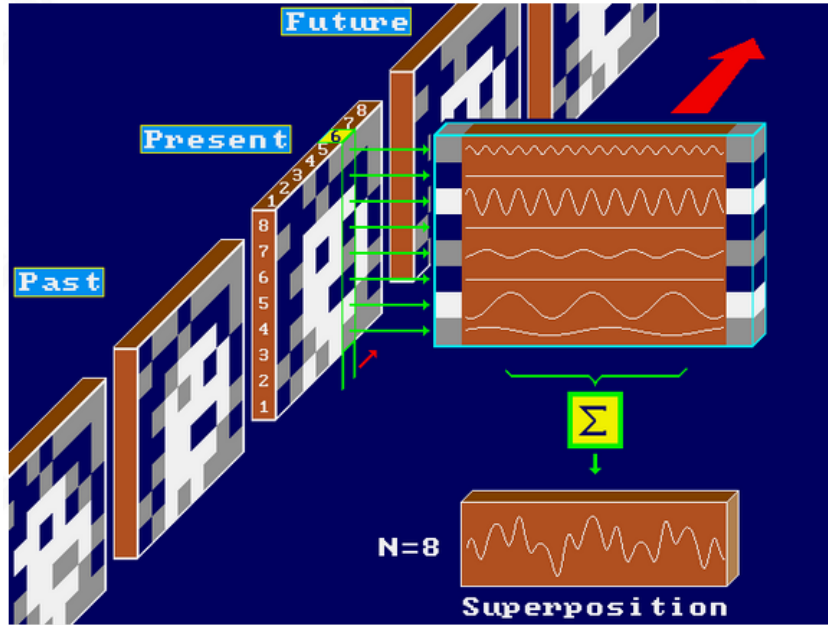
Εφαρμογές στην ιατρική και την επιστήμη

Στο χώρο της ιατρικής, είναι έντονη η δραστηριότητα που σημειώνεται τα τελευταία χρόνια προς την κατεύθυνση της αναπαράστασης του φυσικού περιβάλλοντος χώρου

μέσω ακουστικών ερεθισμάτων ως η μη παρεμβατική/χειρουργική εναλλακτική προσέγγιση σε ασθενείς με ολική τύφλωση. Η παρεμβατική εκδοχή, η οποία απευθύνεται σε ασθενείς με μη αντιστρέψιμες βλάβες του αμφιβληστροειδούς περιλαμβάνει την εμφύτευση φωτοευαίσθητου αισθητήρα ο οποίος μετατρέπει τα φωτεινά ερεθίσματα σε ηλεκτρικά σήματα (όπως συμβαίνει και στη περίπτωση των υγιών ατόμων) με τη βοήθεια ολοκληρωμένου κυκλώματος. Στη συνέχεια τα ηλεκτρικά σήματα διοχετεύονται στο οπτικό νεύρο [9]. Μια άλλη προσέγγιση, η οποία αναπτύχθηκε από τον William H. Dobbelle [10] βρίσκει εφαρμογή σε ασθενείς με μη λειτουργικό οπτικό νεύρο, περιλαμβάνει ταυτόχρονη υποκατάσταση και του οπτικού νεύρου με τεχνητό μόσχευμα το οποίο στη συνέχεια συνδέεται χειρουργικά με τον εγκεφαλικό φλοιό (εγκεφαλικό μόσχευμα) ώστε να γίνεται η μετάδοση των ηλεκτρικών σημάτων μέσω ηλεκτροδίων [11].








Μη παρεμβατικές τεχνικές έχουν επίσης χρησιμοποιηθεί ως υποστηρικτικά μέτρα τόσο σε ασθενείς που πάσχουν από πλήρη τύφλωση όσο και ασθενείς με μερική τύφλωση ή άλλους περιορισμούς στην όραση.

Μια πρόσφατη καινοτομία, η οποία χρησιμοποιήθηκε ως υποστηρικτικό μέτρο σε ασθενείς με πλήρη τύφλωση, αποτελεί -κατ' ουσία- εφαρμογή επαυξημένης πραγματικότητας (augmented reality) και περιλαμβάνει την αισθητηριακή υποκατάσταση της όρασης με ηχητικά σήματα [12]. Ο μηχανισμός περιλαμβάνει μια κάμερα ενσωματωμένη σε ένα ζευγάρι γυαλιά «οράσεως». Η εικόνα που συλλαμβάνεται από την κάμερα μετατρέπεται με τη βοήθεια κατάλληλου λογισμικού σε ηχητικό σήμα με βάση συγκεκριμένες παραδοχές ώστε να σηματοδοτεί προς τον χρήστη της συσκευής τη διαμόρφωση του περιβάλλοντα χώρου [13]. Σταδιακά, ο εγκεφαλικός φλοιός προσαρμόζεται σε αυτού του είδους το ερέθισμα προσομοιώνοντας μερικώς την οπτική αντίληψη του χώρου [14].



Εικόνα 2 Οι εικόνες προσομοιώνονται από πλέγματα εικονοστοιχείων

Μια άλλη υλοποίηση αισθητηριακής υποκατάστασης της όρασης με τη βοήθεια επαυξημένης πραγματικότητας αποτελεί εκείνη που χρησιμοποιείται από τον Neil Harbisson [15]. Ο Harbisson πάσχει από ολική αχρωματοψία και αντιλαμβάνεται τα χρώματα του φυσικού κόσμου μόνο με τη βοήθεια υλισμικού και λογισμικού. Η υποστηρικτική συσκευή αναλαμβάνει να ανιχνεύει τα χρώματα στο περιβάλλον (ακόμη και το υπέρυθρο και το υπεριώδες εκτός του ορατού φάσματος) με τη βοήθεια φορητού αισθητήρα. Στη συνέχεια αυτά μετατρέπονται σε ηχητικά σήματα βάσει “ηχοχρωματικού” πρωτοκόλλου του οποίου μια συνοπτική αναφορά γίνεται στην Εικόνα 3.

PURE SONOCHROMATIC SCALE		
(invisible)	Ultraviolet	Over 717.591 Hz
	Violet	607.542 Hz
	Blue	573.891 Hz
	Cyan	551.154 Hz
	Green	478.394 Hz
	Yellow	462.023 Hz
	Orange	440.195 Hz
	Red	363.797 Hz
(invisible)	Infrared	Below 363.797 Hz

Εικόνα 3 «Ηχοχρωματολόγιο» ενίσχυσης οπτικών δυνατοτήτων

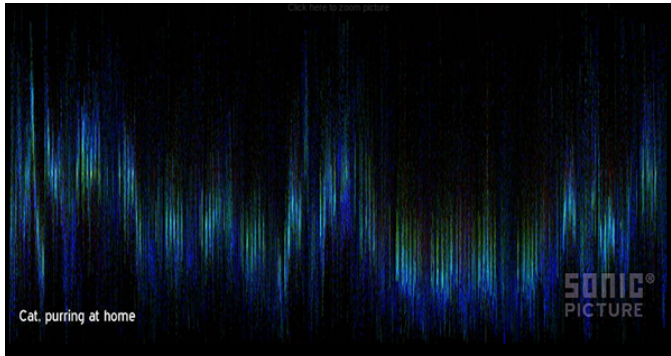
Μεταξύ των λιγότερο “εξωτικών” εφαρμογών της μετατροπής ήχου σε εικόνα αποτελεί η γνωστή συσκευή υπερηχογραφήματος η οποία χρησιμοποιείται στην κλινική εξέταση και που επιτρέπει την απεικόνιση των εσωτερικών μαλακών ιστών του ανθρώπινου σώματος με τη βοήθεια υψίσυχνων (2-18 Mhz) ήχων [16].

Εφαρμογές στην επιστήμη της διασκέδασης και της ψυχαγωγίας

Σε αυτό το πεδίο υπάρχει επίσης πληθώρα εφαρμογών τόσο για λόγους ψυχαγωγίας-διασκέδασης όσο και καλλιτεχνικού ή και επιστημονικού πειραματισμού.

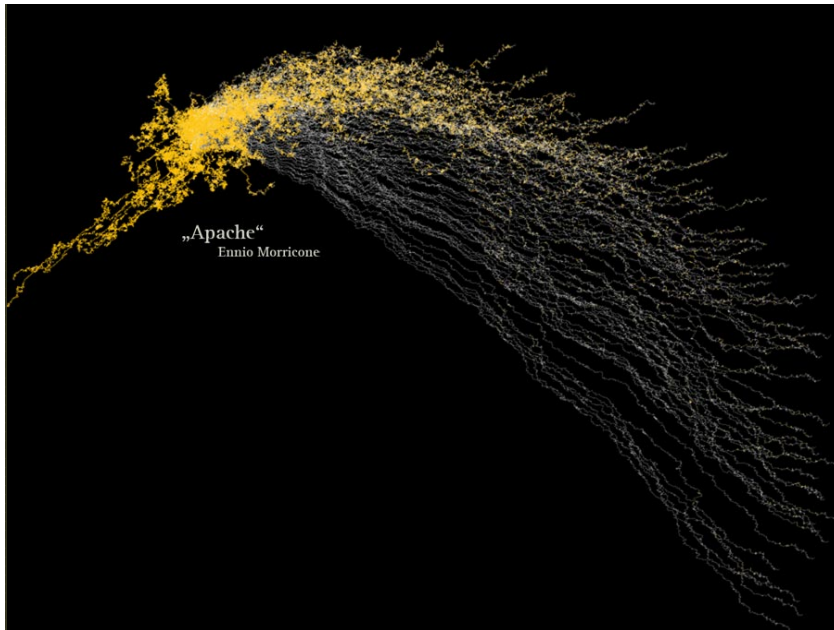
Στον χώρο της καλλιτεχνικής έκφρασης και πειραματισμού συναντάμε και τις δύο κατευθύνσεις μετατροπής της πληροφορίας: εικόνας σε ήχο και ήχου σε εικόνα.

Ο Γερμανός καλλιτέχνης Rainer Tautenhahn συλλαμβάνει ήχους με τη βοήθεια ενός εξαιρετικά ευαίσθητου μικροφώνου και στη συνέχεια αντιστοιχίζει μέσω λογισμικού και με τη βοήθεια αλγορίθμου τις ηχητικές συχνότητες σε συχνότητες του ορατού φάσματος [17] δημιουργώντας εικόνες όπως η Εικόνα 4 χρησιμοποιώντας ως θέματα το ηχητικό τοπίο ενός εμπορικού κέντρου, ενός ζωολογικού κήπου ή ενός παιδικού δωματίου, ήχους ζώων και τις συχνότητες που παράγονται κατά τη μάσηση ενός μήλου.

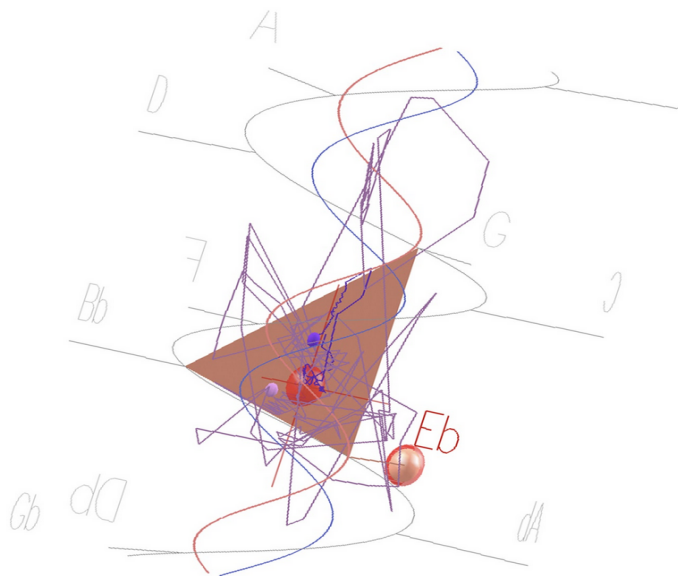


Εικόνα 4 Ηχητικό "πορτραίτο"

Ποικίλης αισθητικής αποτελέσματα προκύπτουν και από τις διάφορες εφαρμογές “οπτικοποίησης” ήχων και κυρίως μουσικής. Στην πιο συνηθισμένη περίπτωση, αυτές οι εφαρμογές προσφέρουν τη δυνατότητα δυναμικής σε πραγματικό χρόνο οπτικοποίησης της μουσικής που αναπαράγεται από την εφαρμογή με τη χρήση αλγορίθμων οι οποίοι αξιοποιούν διάφορα στοιχεία του ήχου (συνήθως συχνότητες, ρυθμό κλπ.). Αυτά αποδίδονται με σχέδια και χρώματα επί της οθόνης. Υπάρχουν όμως και πιο εξελιγμένες τέτοιες εφαρμογές οι οποίες έχουν σχεδιαστεί ώστε να “αντιλαμβάνονται” τη μουσική σύνθεση σε μεγαλύτερο βάθος, σε επίπεδο αρμονίας και συσχετισμού των μουσικών φθόγγων, και να παράγουν περισσότερο εστιασμένες οπτικοποιήσεις. Αυτές οι εφαρμογές παράγουν την αντίστοιχη -δυναμική ή και στατική- οπτικοποίηση είτε σε πραγματικό χρόνο είτε αφ' ότου έχουν επεξεργαστεί ολόκληρη τη σύνθεση [18]. Παρακάτω, στην Εικόνα 5 και Εικόνα 6 δείχνονται απεικονίσεις με τη βοήθεια τέτοιου λογισμικού [19] [20].



Εικόνα 5 Στατική απεικόνιση από την εφαρμογή 'narratives' του Matthias Dittrich



Στον αντίποδα αυτής της πρακτικής βρίσκεται η ιδέα πίσω από τη “Φασματική μουσική”. Η τελευταία αποτελεί μια καλλιτεχνική προσέγγιση βάσει της οποίας η μουσική σύνθεση υπαγορεύεται από φασματογραφικές απεικονίσεις και μαθηματική ανάλυση (λ.χ. θεωρία αριθμών, fractals, αλγόριθμους γενετικής κλπ.) του ηχητικού φάσματος. Κατά την προσέγγιση αυτή, η χροιά του ήχου προκύπτει ως απόρροια της ανάλυσης του ήχου με τη βοήθεια ηλεκτρονικού υπολογιστή και την απεικόνιση των ηχητικών σημάτων [21].

Σκοπός και Δομή

Σκοπός της εργασίας αυτής είναι να πραγματοποιηθεί ο σχεδιασμός και η ανάπτυξη συστήματος λογισμικού το οποίο θα χρησιμοποιεί δεδομένα εικόνας τα οποία θα εξάγει από μια συλλογή φωτογραφιών του χρήστη και στη συνέχεια θα τα χρησιμοποιεί για την αναπαραγωγή ηχητικής επένδυσης παράλληλα με την προβολή ροής εικόνων της συλλογής φωτογραφιών.

Η δομή που ακολουθήθηκε για την εργασία απαρτίζεται από τα κεφάλαια τα οποία αναφέρονται παρακάτω:

Κεφάλαιο 1. Εισαγωγή: Στο κεφάλαιο αυτό γίνεται μια σύντομη αναδρομή σε πρώιμες εκφάνσεις της ιδέας πίσω από το σχεδιασμό του συστήματος λογισμικού αλλά και γενικότερα στη προσπάθεια του ανθρώπου να εφαρμόζει την εικόνα και τον ήχο σε έργα διασκέδασης/ψυχαγωγίας και επιστημονικού ενδιαφέροντος.

Κεφάλαιο 2. Θεωρητικό Υπόβαθρο: στο κεφάλαιο αυτό θέτουμε τη θεωρητική βάση για την κατανόηση της λειτουργίας των αισθήσεων της όρασης και ακοής στον άνθρωπο. Επίσης γίνεται αναφορά στα στοιχεία των αισθήσεων αυτών πάνω στην οποία βασίστηκε η εφαρμογή που αναπτύχθηκε στα πλαίσια της εργασίας. Παρατίθενται επίσης αναφορές στις ιδέες που βρίσκονται πίσω από τις αρχές ανάπτυξης της εφαρμογής αλλά και κάποιες προκλήσεις που σχετίζονται με την ανάπτυξη εφαρμογών σε αυτή την περιοχή. Τέλος γίνεται αναφορά στο σκοπό και τα προσδοκώμενα αποτελέσματα τη εργασίας.

Κεφάλαιο 3. Ανάλυση του συστήματος: Σε αυτό το κεφάλαιο πραγματοποιείται ανάλυση

του συστήματος και αναφέρονται τα κριτήρια με τα οποία επιλέχθηκαν οι χρησιμοποιούμενοι αλγόριθμοι.

Κεφάλαιο 4. Μεθοδολογία ανάπτυξης του συστήματος: γίνεται σύντομη αναφορά στις επικρατούσες μεθοδολογίες ανάπτυξης λογισμικού και παρουσιάζεται η μεθοδολογία που επιλέχθηκε για την ανάπτυξη του συστήματος λογισμικού της παρούσας εργασίας καθώς επίσης και οι λόγοι που οδήγησαν στην επιλογή αυτή.

Κεφάλαιο 5. Σχεδίαση συστήματος: παρατίθενται λεπτομέρειες που αφορούν στη σχεδίαση του συστήματος καθώς και αιτιολόγηση των σχεδιαστικών επιλογών που έγιναν στα πλαίσια της εργασίας ενώ περιγράφεται και η αρχιτεκτονική του συστήματος αλλά και οι εφαρμογές οι οποίες χρησιμοποιήθηκαν ως βάση ανάπτυξης.

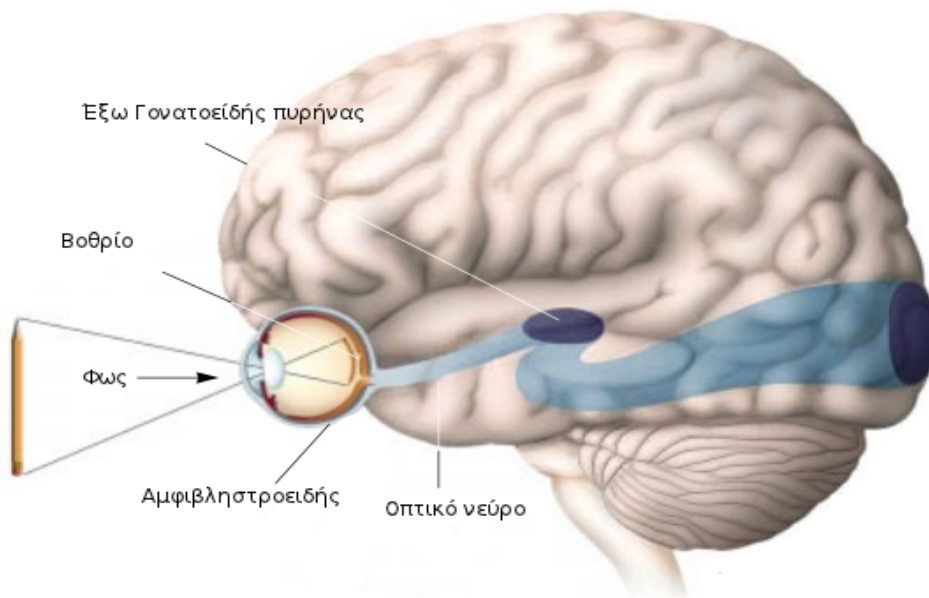
Κεφάλαιο 6. Αποτελέσματα: δίνονται πληροφορίες για τα τελικά αποτελέσματα που προέκυψαν τόσο από την εφαρμογή λογισμικού και τη διαδικασία ανάπτυξής της όσο και από τα χρησιμοποιούμενα εργαλεία και τους διάφορους περιορισμούς οι οποίοι επιβλήθηκαν από αυτά.

Κεφάλαιο 7. Συμπεράσματα και Προοπτικές: παρατίθεται μια σύντομη λίστα με συμπεράσματα που βγήκαν από τη διαδικασία σχεδιασμού και ανάπτυξης και γίνεται μια απόπειρα περιγραφής εκείνων που θεωρούνται πιθανές μελλοντικές προοπτικές ανάπτυξης της εφαρμογής.

2 ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

Οι αισθήσεις της Όρασης και της Ακοής

Ο άνθρωπος μπορεί να αναγνωρίσει τη μορφή οικείου του ατόμου σχεδόν ακαριαία



ιου αυτού.
 υποθετήσει
 εξ αυτές οι
 Πρωτεύον οπ. Ωστόσο
 φλοιός
 συνεργασία
 σύνθετων
 Movshon,
 ήσουμε τα

αισθητήρια όργανα του ανθρώπου ως μικρούς επισημονες οι οποίοι κάνουν υποθέσεις

σχετικά με τον κόσμο” [1]. Στην πραγματικότητα, οτιδήποτε γνωρίζουμε για τον κόσμο γίνεται αντιληπτό μέσω των αισθήσεων. Παραδοσιακά, γίνεται αποδεκτό ότι οι αισθήσεις είναι πέντε: όραση, ακοή, αφή, όσφρηση και γεύση. Τον τελευταίο καιρό έχουν προστεθεί και κάποιες άλλες αισθήσεις στη λίστα αυτή, όπως για παράδειγμα εκείνη του πόνου, της θερμοκρασίας, της μυϊκής τάσης κ.α. Ωστόσο σύμφωνα, με την απλοποιημένη θεώρηση, μπορούμε αυτές να τις εντάξουμε στη γενικότερη αίσθηση της αφής.

Οι ακτίνες του φωτός όταν ανακλαστούν από κάποιο αντικείμενο, για παράδειγμα ένα μολύβι, εισέρχονται στο μάτι και διαπερνούν το φακό. Ο φακός προβάλλει ένα αντεστραμμένο είδωλο πάνω στον αμφιβληστροειδή στο πίσω μέρος του οφθαλμού. Ηλεκτρικά σήματα τα οποία παράγονται από τα ραβδία και τα κωνία, τα εξειδικευμένα κύτταρα-υποδοχείς του ματιού, δρομολογούνται μέσω του οπτικού νεύρου προς τον εγκέφαλο μέχρι τον Έξω Γονατοειδή Πυρήνα από όπου και αναμεταδίδονται. Τμήμα της πληροφορίας η οποία μεταδίδεται στο οπτικό νεύρο φθάνει στον πρωτεύοντα οπτικό φλοιό από όπου διαχέονται σε υψηλότερα τμήματα του φλοιού τα οποία επεξεργάζονται τα γενικά χαρακτηριστικά της εικόνας του μολυβιού όπως το σχήμα, το χρώμα ή την κίνηση.

Παρακάτω παρατίθενται σε συντομία τα βασικά στοιχεία όρασης [2].

Η φωτεινότητα του πεδίου

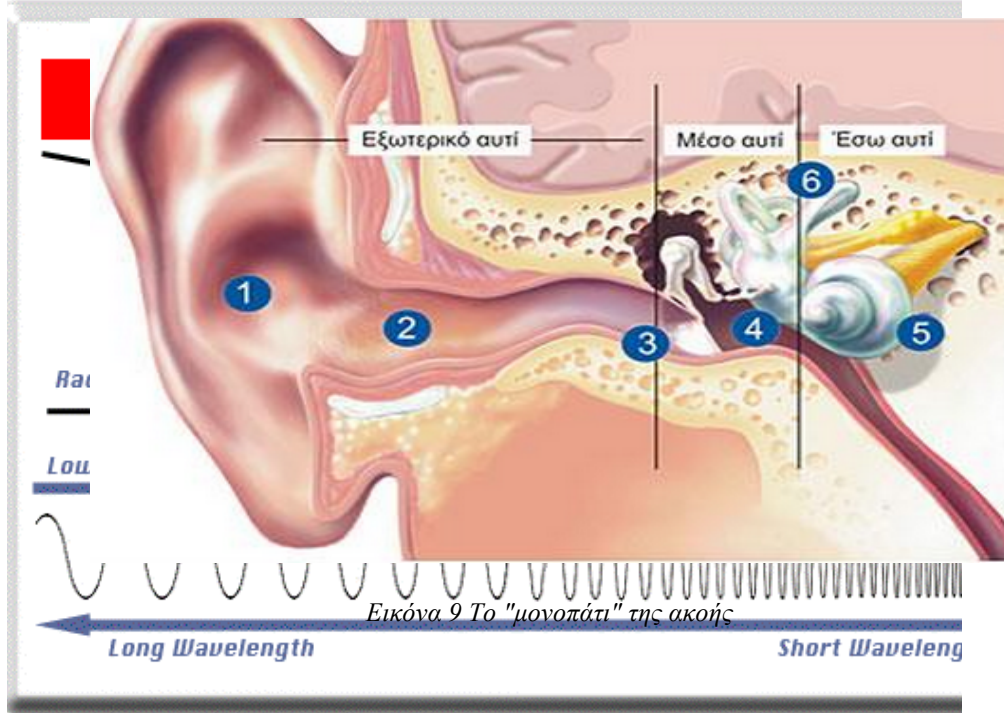
Η φωτεινότητα του πεδίου αλλάζει το επίπεδο της προσαρμογής του αμφιβληστροειδούς. Χαμηλά επίπεδα φωτισμού ενεργοποιούν τα ραβδία. Η πυκνότητα των ραβδίων είναι χαμηλή και επομένως και η οπτική οξύτητα είναι χαμηλή. Αντίθετα, όταν τα επίπεδα φωτισμού είναι υψηλά ενεργοποιούνται τα κωνία και μεγιστοποιείται η οπτική οξύτητα.

Αντίθεση

Το ανθρώπινο μάτι αντιλαμβάνεται αντικείμενα αντιδρώντας στα διαφορετικά επίπεδα φωτισμού στα όρια του οπτικού στόχου. Αυτή η διαφορά στα επίπεδα φωτισμού ονομάζεται αντίθεση.

Χρωματική αντίληψη

Τα αντικείμενα ανακλούν διαφορετικά μήκη κύματος του φωτός και η αντίληψη τους από τον άνθρωπο δημιουργεί την αίσθηση των χρωμάτων. Η αντίληψη των χρωμάτων είναι λειτουργία των κωνίων. Διαφορετικές συχνότητες του ορατού φάσματος διεγείρουν διαφορετικούς πληθυσμούς από κωνία. Με αυτόν τον τρόπο μπορούν να παραχθούν όλα τα χρώματα του φάσματος. Η όραση μας, πέρα από την φασματική ευαισθησία, είναι σε θέση να διακρίνει τη λαμπρότητα (φωτεινότητα) του χρώματος αλλά και τον κορεσμό (το



μαγνητικού φάσματος και
κάτω εικόνα:

Συνεπώς το ερυθρό χρώμα αντιστοιχεί στη χαμηλότερη συχνότητα του ορατού φάσματος ενώ το μωβ στην υψηλότερη.

Το έξω αυτί συλλέγει τον ήχο στην κόγχη και τον μεταφέρει κατά μήκος του ακουστικού πόρου στο μέσο αυτί. Εκεί οι ήχοι δονούν την τυμπανική μεμβράνη του αυτιού. Αυτή, με την σειρά του, κινεί διαδοχικά τα τρία μικρά οστά (σφύρα, άκμονας και αναβολέας) που βρίσκονται στο μέσο αυτί. Η κίνηση αυτή προκαλεί δονήσεις στον κοχλία του εσωτερικού αυτιού. Μέσω των εξειδικευμένων τριχοφόρων κυττάρων, οι δονήσεις μετατρέπονται σε νευρικές ώσεις (ηλεκτρικά σήματα) τα οποία στη συνέχεια μεταδίδονται από το κοχλιακό τμήμα του ακουστικού νεύρου στον ακουστικό φλοιό για αναγνώριση [3].

Κάθε σώμα που τίθεται σε ταλάντωση με την επίδραση κάποια δύναμης, προκαλεί κίνηση των μορίων του ελαστικού μέσου που το περιβάλλει. Το ταλαντευόμενο σώμα τότε αποτελεί ηχητική πηγή και το ελαστικό μέσο μπορεί να βρίσκεται σε αέρια, υγρή ή στερεή κατάσταση. Η ταλάντωση της ηχητικής πηγής διαδίδεται στο περιβάλλον λόγω της μεταβίβασης ενέργειας μέσω της ταλάντωσης των μορίων. Η ταλάντωση αυτή των μορίων του περιβάλλοντος της ηχητικής πηγής δημιουργεί το ηχητικό κύμα. Όταν ένα μόριο του μέσου ενεργοποιηθεί από την ηχητική πηγή, εκτελεί πλήρη ταλάντωση μέχρι να επανέλθει στην αρχική του θέση. Η κίνηση αυτή αποτελεί μια απλή αρμονική κίνηση ή έναν κύκλο ο οποίος παριστάνεται γραφικά με μια ημιτονοειδή καμπύλη. Όσο διαρκεί η πηγή να βρίσκεται σε ταλάντωση, τα μόρια του μέσου ενεργοποιούνται συνεχώς κάνοντας αρμονικές κινήσεις. Για να παραχθεί λοιπόν και να φτάσει στο αυτί ένας ήχος απαιτούνται:

- Μια πηγή
- Μια δύναμη που θα θέσει την πηγή σε ταλάντωση
- Ένα υλικό μέσο μεταφοράς των παραγόμενων ηχητικών κυμάτων λόγω της ταλάντωσης, από την ηχητική πηγή στο αυτί

Ήχος και Ακουστικό ερέθισμα

Συχνότητα του ήχου

Η συχνότητα του ηχητικού κύματος είναι ο αριθμός των κύκλων που παράγονται σε διάστημα ενός δευτερολέπτου, Επομένως η μονάδα μέτρησης της συχνότητας είναι

κύκλος/δευτερόλεπτο ή cycle/second ή cps ή Hz (Hertz)

Ένταση του Ήχου

Η ένταση του ήχου αναφέρεται στην ισχύ του ταλαντούμενου μορίου του μέσου διάδοσης του ήχου, ή στο ποσό της ενέργειας που ρέει δια μέσο του περιβάλλοντος την ηχητική πηγή υλικού. Η απευθείας μέτρηση της έντασης του ήχου είναι δύσκολη. Όμως, μιας και ο ήχος συνίσταται από συνεχείς διακυμάνσεις της πίεσής του, είναι πιο πρακτικό να αποδώσουμε την ένταση του ήχου σε όρους διακύμανσης της πίεσής του. Η σχέση αυτή είναι η παρακάτω:

$$I = p^2/(pc),$$

όπου I: η ένταση, p: η πίεση, c: η ταχύτητα του φωτός

Decibel

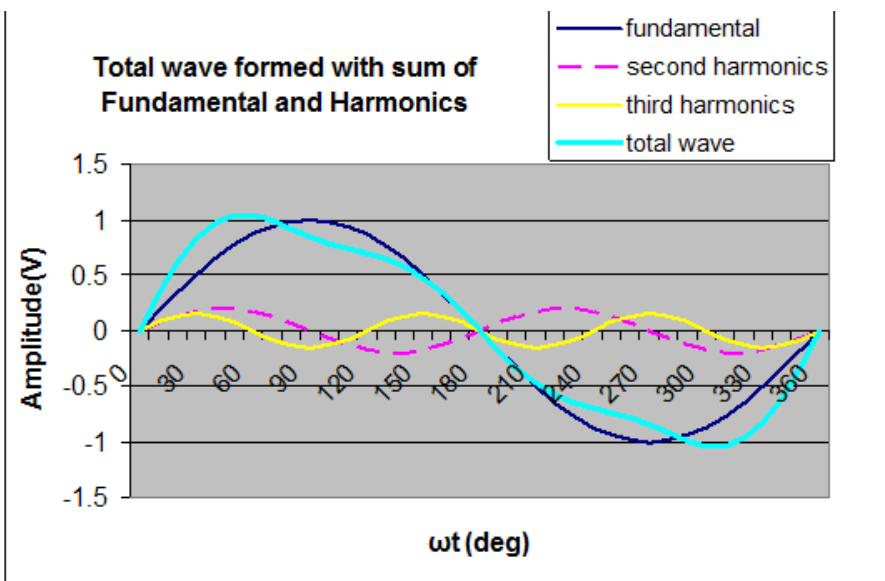
Πολλές υποκειμενικές αισθήσεις είναι ανάλογες με τη λογαριθμική λειτουργία του φυσικού ερεθίσματος που τις παράγει. Για την απόκριση του ανθρώπινου αυτιού στο ηχητικό ερέθισμα επινοήθηκε η μονάδα μέτρησης decibel (db). Ο γενικός τύπος που χρησιμοποιείται είναι:

$$1 \text{ db} = 10 \log(I_1/I_0) = 10 \log(p_1^2/p_0^2) = 20 \log(p_1/p_0)$$

όπου τα I_0 και p_0 αναφέρονται στις τιμές κατωφλίου της έντασης και της πίεσης αντιστοίχως όπου ο ήχος γίνεται ακουστός από το ανθρώπινο αυτί.

Φάσμα του Ήχου

Η απλούστερη μορφή του ήχου είναι ο καθαρός τόνος. Ο ήχος αυτός μπορεί να περιγραφεί με επάρκεια με τη συχνότητα και την ένταση του. Η παρουσία όμως καθαρών ήχων στο περιβάλλον είναι αρκετά σπάνια. Οι περισσότεροι ήχοι στη φύση είναι σύνθετοι και αποτελούνται από ήχους διαφόρων συχνοτήτων και εντάσεων. Οι μουσικοί ήχοι χαρακτηρίζονται από περιοδικότητα, δηλαδή η κυματομορφή τους επαναλαμβάνεται με τη συχνότητα του πιο χαμηλόσυχνου συνθετικού τους. Η συχνότητα αυτή ονομάζεται θεμελιώδης. Εκτός από τη θεμελιώδη, ένα σύνθετο κύμα



αν είναι πολλαπλάσια της
 θετικά που έχουν δύο ή τρεις
 ιχα δεύτερη ή τρίτη αρμονική

Εικόνα 10 Κύμα με δύο αρμονικές

Όταν δύο καθαροί τόνοι έχουν την ίδια συχνότητα και ένταση αλλά διαφέρουν στο χρόνο μεταξύ τους κατά μισή περίοδο, λέμε ότι “τα κύματα βρίσκονται σε αντίθετη φάση” και αλληλοεξουδετερώνονται.

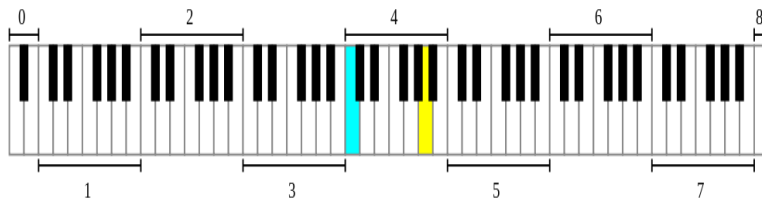
Οι αρμονικές ενός ήχου συνήθως προκαλούνται από την ηχητική πηγή που παράγει και τη θεμελιώδη συχνότητα. Στη φύση συναντάμε ήχους οι οποίοι αποτελούνται από περισσότερους του ενός τόνου οι οποίοι είναι ως επι το πλείστον αρμονικοί μεταξύ τους. Οι ήχοι αυτοί ονομάζονται φθόγγοι. Τέλος οι σύνθετοι ήχοι από τους οποίους απουσιάζει η αρμονία και αποτελούνται από διάφορες συχνότητες, εντάσεις και φάσεις ονομάζονται θόρυβος.

Στοιχεία Ψυχοακουστικής

Αντικείμενο της ψυχοακουστικής είναι ο καθορισμός των σχέσεων μεταξύ των φυσικών παραμέτρων τους ακουστικό ερεθίσματος και των προκαλουμένων από αυτό ψυχολογικών (υποκειμενικών) αισθήσεων.

Ύψος

Το φυσικό μέγεθος της συχνότητας ενός ακουστού ήχου είναι πρωταρχικά υπεύθυνο για την υποκειμενική αντίληψη ή ψυχολογική αίσθηση του ύψους. Όταν αυξάνει η συχνότητα ενός ήχου, το ύψος του γίνεται “οξύτερο”. Χαμηλής συχνότητας ήχοι σχετίζονται με το “χαμηλό” ύψος. Οποσδήποτε όμως η αίσθηση του ύψους δεν ακολουθεί γραμμικά το φυσικό μέγεθος της συχνότητας του ήχου. Ως μονάδα μέτρησης του ύψους καθορίστηκε το mel. Τα 1000 mel καθορίστηκαν ως το ύψος ενός τόνου 1000 Hz σε ένταση 40 db επιπέδου πίεσης ήχου. Τόνοι που απέχουν μεταξύ τους κατά ίσο αριθμό mels, θα εμφανίζουν και την ίδια διαφορά ύψους. Με διάφορες ψυχοφυσικές μεθόδους καθορίστηκαν οι διαβαθμίσεις της συχνότητας ενός τόνου, έτσι ώστε το ύψος του να είναι διπλάσιο ή υποδιπλάσιο σε σχέση με το ύψος τόνου αναφοράς. Όταν η συχνότητα ενός τόνου διπλασιάζεται (ενώ η ένταση παραμένει σταθερή), το ύψος του τόνου αυξάνει κατά μια οκτάβα, ενώ όταν η συχνότητα υποδιπλασιαστεί, το ύψος χαμηλώνει κατά μια οκτάβα [3]. Παρακάτω φαίνεται η αντιστοιχία μεταξύ της απόλυτης συχνότητας για τις νότες ενός πιάνου 88 πλήκτρων. Με κυανό και κίτρινο χρώμα έχουν επισημανθεί η “μέση Ντο” και η “μέση Λα” αντιστοίχως [4].



Key number	Helmholtz name	Scientific name	Frequency (Hz)
88	c ⁸⁸ 5-line octave	C8 Eighth octave	4186.01
87	b ⁸⁷	B7	3951.07
86	a ⁸⁶ /b ⁸⁶	A ⁷ /B ⁷	3729.31
85	a ⁸⁵	A7	3520.00
84	g ⁸⁴ /a ⁸⁴	G ⁷ /A ⁷	3322.44
83	g ⁸³	G7	3135.96
82	f ⁸² /g ⁸²	F ⁷ /G ⁷	2959.96
81	f ⁸¹	F7	2793.83
...
51	b ⁵¹	B4	493.883
50	a ⁵⁰ /b ⁵⁰	A ⁴ /B ⁴	466.164
49	a ⁴⁹	A4 A440	440.000
48	g ⁴⁸ /a ⁴⁸	G ⁴ /A ⁴	415.305
47	g ⁴⁷	G4	391.995
46	f ⁴⁶ /g ⁴⁶	F ⁴ /G ⁴	369.994
45	f ⁴⁵	F4	349.228
44	e ⁴⁴	E4	329.628
43	d ⁴³ /e ⁴³	D ⁴ /E ⁴	311.127
42	d ⁴²	D4	293.665
41	c ⁴¹ /d ⁴¹	C ⁴ /D ⁴	277.183
40	c ⁴⁰ 1-line octave	C4 Middle C	261.626
...
2	A ₂ /B ₂	A ⁰ /B ⁰	29.1352
1	A ₁ sub-contra-octave	A0 Double Pedal A	27.5000

Εικόνα 11 Αντιστοιχία νότας-Hz σε νότες πιάνου 88 πλήκτρων

Διάκριση των διαφορών στη Συχνότητα

Η μικρότερη διακύμανση στη συχνότητα που γίνεται αντιληπτή από το ανθρώπινο αυτί σαν αλλαγή του ύψους του τόνου, για συχνότητες κάτω των 1000 Hz είναι της τάξης του 1-3 Hz. Για τόνους ψηλότερων συχνοτήτων, απαιτείται μια προοδευτικά αυξανόμενη διακύμανση της συχνότητας προκειμένου η διαφορά του ύψους να γίνει αντιληπτή. Η διακύμανση αυτή έχει μια σταθερή αναλογία σε σχέση με τη συχνότητα του τόνου η οποία είναι περίπου 0,002 της συχνότητας του τόνου [3]. Επομένως όταν βρισκόμαστε, για παράδειγμα, στην περιοχή των 4000 Hz, απαιτούνται διακυμάνσεις περίπου 8 Hz ώστε να γίνουν αντιληπτές.

Ακουστότητα

Ακουστότητα στην ψυχοακουστική θεωρείται η ψυχολογική αίσθηση του φυσικού μεγέθους της έντασης του ηχητικού ερεθίσματος. Η ακουστότητα αν και συσχετίζεται με την ένταση του ακουστικού ερεθίσματος που την προκαλεί, ωστόσο η σχέση δεν είναι γραμμική: για να φθάσει ένας ήχος στο κατώφλι της ακουστότητας απαιτείται μεγαλύτερη ένταση στις χαμηλές και τις ψηλές συχνότητες από ότι στις μεσαίες.

Χροιά

Το φάσμα ενός ήχου σχετίζεται με την ψυχολογική αίσθηση της χροιάς. Όταν διαφέρει η κυματομορφή και σε επέκταση το φάσμα δύο ή και περισσότερων ήχων, οπωσδήποτε οι ήχοι αυτοί έχουν διαφορετική χροιά. Το ανθρώπινο αυτί, για παράδειγμα, είναι σε θέση να αναγνωρίσει τη διαφορά μεταξύ π.χ. ενός σαξοφώνου και μιας τρομπέτας που παίζουν την ίδια νότα εξαιτίας των διαφορών στο φάσμα του κάθε οργάνου. Το φάσμα αυτό είναι προϊόν τόσο της πολυπλοκότητας των δονήσεων της ηχητικής πηγής όσο και των ειδικών χαρακτηριστικών αντηχείων που φέρει η πηγή.

Αντίληψη του Περιβάλλοντος μέσω της Εικόνας και μέσω του Ήχου

Η έμφαση που δίδεται στον Δυτικό πολιτισμό σε όσα γίνονται αντιληπτά μέσω της αίσθησης της όρασης, κάνει δύσκολο το να αντιληφθεί κάποιος -εάν δεν έχει προβλήματα όρασης- ότι ο κόσμος της ακοής είναι ένας κόσμος συμβάντων ενώ ο κόσμος της όρασης είναι ένας κόσμος αντικειμένων (Walter J. Ong) [6]. Προβαλλόμενη στη σχέση μεταξύ του ήχου και της εικόνας, αυτή η έμφαση, οδηγεί στην παρατήρηση ότι στην κουλτούρα μας επικρατεί η οπτική απεικόνιση έναντι της ηχητικής. Ωστόσο η πρώτη αίσθηση που χρησιμοποιείται από τον άνθρωπο κατά την εμβρυική ηλικία είναι εκείνη της ακοής. Οι όραση και η ακοή έχουν κοινούς αισθητηριακούς τόπους αλλά και ειδοποιούς διαφορές. Επίσης η αντιληπτικότητα του κόσμου μέσω της κάθε μιας από αυτές τις αισθήσεις παρουσιάζει διαφορές.

Ομοιότητες:

- Συμβάλλουν αμφότερες στην ολοκληρωμένη αντίληψη του περιβάλλοντος κόσμου
- Σε βιοχημικό/βιοηλεκτρικό επίπεδο τόσο τα οπτικά όσο και τα ακουστικά ερεθίσματα μετατρέπονται τελικά σε ηλεκτρικές ώσεις οι οποίες καταλήγουν και αναλύονται από τα ανώτερα επίπεδα του εγκεφαλικού φλοιού.
- Επηρεάζονται από τη φύση του υλικού που περιβάλλει τα αντίστοιχα αισθητήρια όργανα. Για παράδειγμα τόσο η όραση όσο και η ακοή επηρεάζονται όταν η παρατήρηση/ακρόαση γίνεται μέσα σε υδάτινο περιβάλλον
- Συντελούν στη βιωσιμότητα του ατόμου

Διαφορές:

- Αυτό που γίνεται αντιληπτό μέσω της όρασης, αν εξαιρεθούν οι σκιές και οι αντανακλάσεις, αντιστοιχεί ακριβώς σε αυτό προς το οποίο κοιτάζει κάποιος. Στην περίπτωση της ακοής, η χωρική ακρίβεια δεν είναι τόσο δεδομένη: ο ακροατής κάποιου ήχου αντιλαμβάνεται από που έρχεται ο ήχος αλλά όχι απαραίτητα και ποια είναι η πρωτογενής προέλευσή του [7].
- Επίσης δεδομένη δεν είναι ούτε η χρονική ακρίβεια μεταξύ του ηχητικού ερεθίσματος σε σχέση με το οπτικό. Η αντίληψη ενός ήχου συχνά προηγείται της οπτικής επαφής με την πηγή που τον παράγει. Συχνά δεν είμαστε σε θέση να δούμε εκείνο που ακούμε (και το αντίστροφο).
- Η οπτική αρμονία είναι κάτι που μπορεί να απαντηθεί τυχαία στη φύση. Για παράδειγμα ένα πολύχρωμο αισθητικά αρμονικό τοπίο. Κάτι τέτοιο δεν φαίνεται να ισχύει για τους ήχους. Οι φυσικοί ήχοι είναι κατά κανόνα μη-αρμονικοί [7]. “Εκτός ίσως από το κελάηδισμα κάποιων πτηνών, ο άνθρωπος δεν θα ερχόταν σε επαφή με τη μουσική εάν δεν την είχε εφεύρει” (Claude Lévi-Strauss).

Δημοφιλείς τύποι αρχείων εικόνας και το πρόβλημα της μετατροπής δεδομένων εικόνας σε στοιχεία Ήχου

Όπως είδαμε παραπάνω η απόδοση των δεδομένων εικόνας σε ήχο, όταν γίνεται για ιατρικούς, επιστημονικούς λόγους, κατά την εφαρμογή τους για παράδειγμα σε

υποστηρικτικές της όρασης συσκευές, απαιτείται αυστηρά να παράγουν ήχους οι οποίοι να δίνουν καλή ηχητική αναπαράσταση του περιβάλλοντα χώρου. Για παράδειγμα χρειάζεται να τροποποιούν την ένταση ή την συχνότητα του ήχου για να αποδώσουν την απόσταση του χρήστη από τα αντικείμενα αλλά και την καθ' ύψος τοποθέτηση των αντικειμένων. Δεν απαιτείται ωστόσο να υπακούει η εφαρμογή αυτή σε κανόνες ακουστικής αισθητικής. Η αρμονία των παραγόμενων ήχων δεν είναι αυτοσκοπός και οι μόνοι “αισθητικής φύσεως” περιορισμοί που μπαίνουν για τους ήχους αυτούς υπαγορεύονται από την απαίτηση να είναι ευνόητοι και καταληπτοί για τον χρήστη και να κυμαίνονται σε καλώς ανεκτές συχνότητες και εντάσεις μιας και η έκθεση του χρήστη σε αυτές μπορεί να είναι παρατεταμένη.

Οι εφαρμογές που κινούνται στο χώρο της καλλιτεχνικής έκφρασης και πειραματισμού, από την άλλη μεριά, δεν χρειάζεται να αποδίδουν με σαφήνεια τον δισδιάστατο ή τον τρισδιάστατο χώρο ούτε αξιώνουν ρόλους χρήσης με υψηλό βαθμό κρισιμότητας. Για αυτές τις εφαρμογές, επομένως, η “πιστότητα” στο απεικονιζόμενο θέμα δεν χρειάζεται να είναι αυστηρή και οι απαιτήσεις εστιάζονται περισσότερο στο αισθητικό κομμάτι και στην καλλιτεχνική διάσταση της ηχητικής αξιοποίησης των δεδομένων εικόνας.

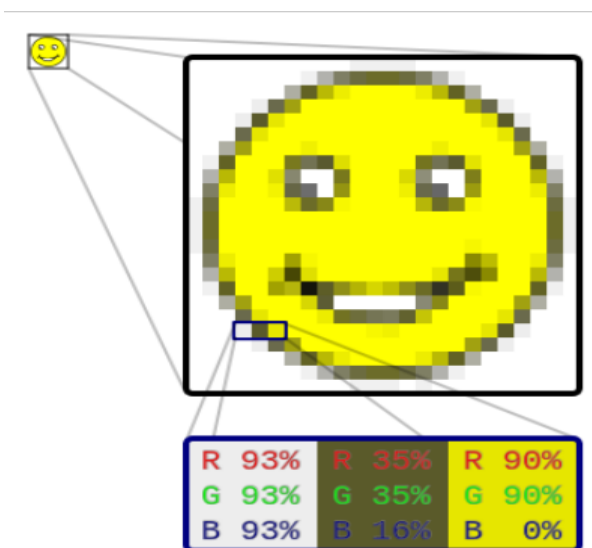
Στα πλαίσια της παρούσας Διπλωματικής εργασίας θα εστιάσουμε στην περίπτωση της ηχητικής “απόδοσης” των δεδομένων που εξάγονται από μια εικόνα δύο διαστάσεων.

Απλά Δεδομένα Εικόνας

Οι εικόνες δύο διαστάσεων μπορούν να αποθηκευτούν σε ψηφιακό αποθηκευτικό μέσο υπό τη μορφή διαφόρων μορφών (format) αρχείων. Συνοπτικά, τα format αυτά διακρίνονται κυρίως με βάση τον τρόπο με τον οποίο αποθηκεύεται η πληροφορία: ασυμπιεστη, συμπίεσμένη ή διανυσματική μορφή. Για να είναι δυνατή η απεικόνιση μιας φωτογραφίας σε έναν υπολογιστή ή σε έναν εκτυπωτή, είναι απαραίτητη η απόδοση της φωτογραφίας υπό τη μορφή πλέγματος (raster), δηλαδή από μια ομάδα εικονοστοιχείων (pixels) κάθε ένα εκ των οποίων αντιπροσωπεύεται από έναν αριθμό δυαδικών ψηφίων (bits) τα οποία αντιστοιχούν στο χρώμα του στοιχείου. Για τις εικόνες υπό τη μορφή πλέγματος, το μέγεθος του αρχείου είναι ανάλογο προς τον αριθμό εικονοστοιχείων που αποτελούν την εικόνα και το βάθος χρώματος. Το ακριβές μέγεθος του αντίστοιχου

αρχείου καθορίζεται από τον αλγόριθμο συμπίεσης που χρησιμοποιείται σε κάθε περίπτωση. Για παράδειγμα ο αλγόριθμος συμπίεσης που χρησιμοποιείται στα αρχεία png (Portable Network Graphics) αποτελεί συνδυασμό των αλγορίθμων LZ77 [22] και της κωδικοποίησης Huffman [23] και παρέχει κωδικοποίηση χωρίς απώλειες ενώ τα δημοφιλή αρχεία JFIF/JPEG κωδικοποιούνται με τη μέθοδο συμπίεσης jpg (Joint Photographic Experts Group) η οποία οδηγεί σε απώλεια πληροφορίας κατά την κωδικοποίηση. Οι διανυσματικές (vector) εικόνες, σε αντίθεση με τις εικόνες τύπου raster η πληροφορία που αποθηκεύεται στα αντίστοιχα αρχεία αποτελεί μια “γεωμετρική περιγραφή” της εικόνας η οποία μπορεί να απεικονιστεί ομαλά σε κάθε επιθυμητό μέγεθος και επομένως το μέγεθος απεικόνισης της εικόνας δεν επηρεάζει το μέγεθός της. Τέτοιου τύπου κωδικοποίηση αποτελεί η SVG (Scalable Vector Graphics) [24]. Ενώ υπάρχουν και τεχνικές που συνδυάζουν τις δύο παραπάνω, για παράδειγμα το PDF (Portable Document Format) [25].

Στα πλαίσια της παρούσας διατριβής θα περιοριστούμε σε εικόνες διαθέσιμες σε μορφή πλέγματος. Σε αυτή τη μορφή η εικόνα αντιπροσωπεύεται από ένα πίνακα ψηφίων (bitmap) όπως φαίνεται στην Εικόνα 12 όπου το κάθε εικονοστοιχείο στη μεγέθυνση εμφανίζεται ως τετράγωνο πλακίδιο. Στη συνέχεια, το κάθε εικονοστοιχείο αναλύεται στα χρώματα κόκκινο, πράσινο και μπλε των οποίων το ποσοστό καθορίζει το τελικό χρώμα του εικονοστοιχείου.



Ζητήματα κατά την αξιοποίηση δεδομένων εικόνας

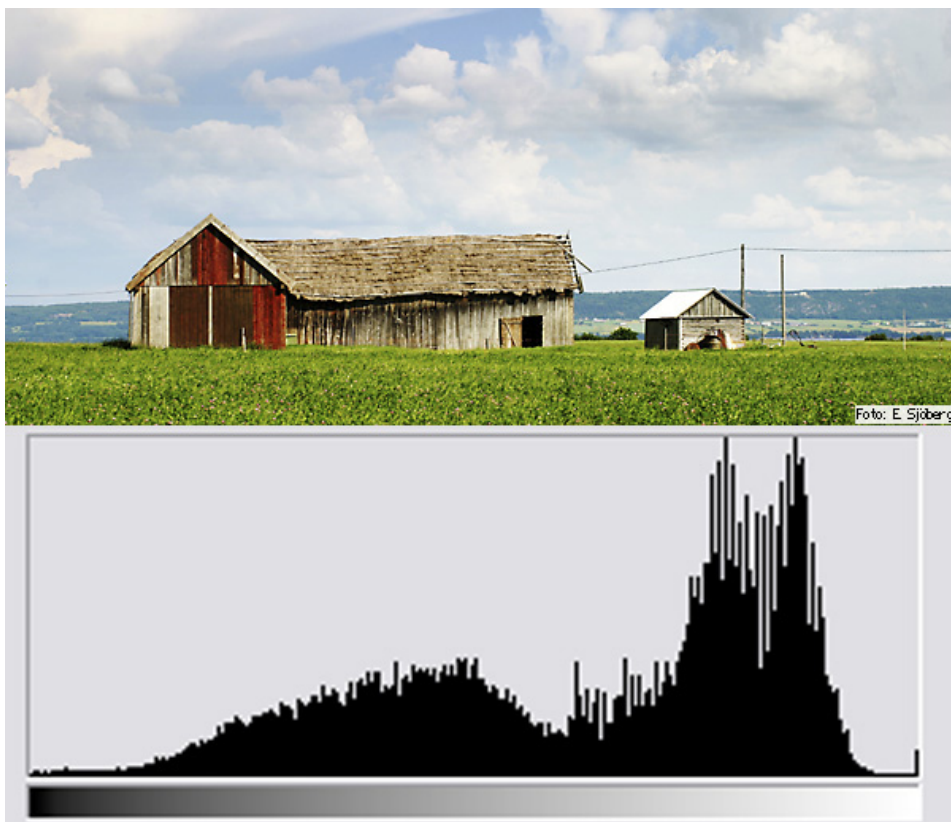
Τίθενται τα παρακάτω γενικά ζητήματα όσον αφορά την αξιοποίηση των δεδομένων εικόνας για την παραγωγή ηχητικής επένδυσης για ροή στατικών εικόνων στα πλαίσια του πειραματισμού με αξιώσεις αισθητικού αποτελέσματος:

- Σε αντίθεση με την κινούμενη εικόνα (λ.χ. κινηματογράφος) ο θεατής μιας ροής στατικών εικόνων αντιλαμβάνεται τη στατικότητα της απεικόνισης. Συνεπώς η ηχητική επένδυση δε μπορεί να υπαγορευθεί από το ρυθμό με τον οποίο εξελίσσεται η δράση. Στην περίπτωση του κινηματογραφικού έργου, για παράδειγμα, μια έντονη σκηνή καταδίωξης θα συνδυαζόταν πιθανότατα με κάποια ζωηρή μελωδία σε γρήγορο ρυθμό. Στην περίπτωσή μας δεν υπάρχει προφανής αντιστοίχιση του προβαλλόμενου θέματος με τα γενικά στοιχεία της ηχητικής επένδυσης οπότε κάτι τέτοιο δε μπορεί να αξιοποιηθεί για την παραγωγή της ηχητικής επένδυσης.
- Ο τρόπος με τον οποίο ο θεατής θα προσπελάσει οπτικά την κάθε εικόνα -για το επίπεδο της παρούσας εργασίας- δε μπορεί να προβλεφθεί αλλά ούτε και να μεθοδευτεί εκ των προτέρων. Είναι πιθανό να εστιάσει σε κάποιο εξέχον στοιχείο της εικόνας, στη συνέχεια να “περιπλανηθεί” σε άλλες λιγότερο “ενδιαφέρουσες” περιοχές της και στη συνέχεια να επιστρέψει στο αρχικό σημείο εστίασης. Και αυτή η οπτική περιπλάνηση μπορεί να διαρκέσει μερικά κλάσματα του δευτερολέπτου. Επομένως, ακόμη και εάν είχαμε την τεχνική δυνατότητα να γνωρίζουμε σε ποιο σημείο της εικόνας εστιάζει κάθε φορά ο θεατής, θα ήταν δύσκολο να προσαρμόσουμε την ηχητική επένδυση με ομαλό τρόπο σε αυτό το στοιχείο.
- Η τάξη της “πυκνότητας” της πληροφορίας σε μια εικόνα σε επίπεδο εικονοστοιχείων κυμαίνεται από εκατοντάδες χιλιάδες έως μερικά εκατομμύρια εικονοστοιχεία ανά εικόνα ενώ η παραμονή της κάθε εικόνας στην οθόνη κατά τη

ροή εικόνων συνήθως δεν έχει διάρκεια μεγαλύτερη από μερικά δευτερόλεπτα. Επομένως δε μπορούμε να έχουμε πυκνότητα ακουστικής πληροφορίας που να αντιστοιχεί στο σύνολο των εικονοστοιχείων χωρίς το τελικό αποτέλεσμα να μην είναι θόρυβος.

- Τα επιμέρους στοιχεία μια εικόνας, όπως τα χρώματα, η φωτεινότητα, η αντίθεση αλλά και τα απεικονιζόμενα γεωμετρικά στοιχεία δεν έχουν προφανή αντιστοίχιση με ηχητικά στοιχεία όσον αφορά το αισθητικό επίπεδο του αποτελέσματος. Για παράδειγμα δεν υπάρχει μονόσημη απάντηση στην ερώτηση: “ποιος ήχος αποτελεί κατάλληλη επένδυση για τη φωτογραφία ενός μπλε μολυβιού σε κόκκινο φόντο”.
- Η επεξεργασία της εικόνας ώστε να εξαχθούν τα δεδομένα ήχου πρέπει να γίνεται σε πραγματικό χρόνο και να χρησιμοποιεί εύλογη ποσότητα υπολογιστικών πόρων.

Υπάρχουσες εφαρμογές οι οποίες παράγουν ήχο χρησιμοποιώντας δεδομένα εικόνας συχνά αξιοποιούν ως είσοδο για την παραγωγή ήχου την ιστογραμματική ανάλυση της εικόνας.



Εικόνα 13 Ιστόγραμμα

Ενώ κάποιες άλλες, προσομοιώνουν την οριζόντια προσπέλαση της εικόνας και χρησιμοποιούν τα επίπεδα του κόκκινου-πράσινου-μπλε ως είσοδο για την ηχητική επένδυση. Φυσικά οι υλοποιήσεις ποικίλουν και δεν εξαντλούνται σε αυτές.

Σκοποί και προσδοκώμενα αποτελέσματα

Γενικοί σκοποί

Η παρούσα εργασία επιδιώκει να παρουσιάσει μια υλοποίηση ηχητικής επένδυσης στατικής ροής εικόνων σε πραγματικό χρόνο. Διαρθρώνεται με βάση τους παρακάτω σκοπούς:

- Την εύρεση και υλοποίηση αλγορίθμου/αλγορίθμων ώστε να εξάγονται σε

πραγματικό χρόνο δεδομένα από στατική εικόνα τα οποία να μπορούν να αξιοποιηθούν για την παραγωγή ήχου

- Την αξιοποίηση των εξαχθέντων δεδομένων για την παραγωγή ηχητικής επένδυσης
- Την ανάλυση, σχεδίαση και υλοποίηση έργου λογισμικού που θα συνιστά ένα ολοκληρωμένο σύστημα για την προβολή ροής εικόνων με παράλληλη παραγωγή ηχητικής επένδυσης

Οι επιμέρους στόχοι

Η παρούσα εργασία θα συνεισφέρει και στους παρακάτω στόχους:

- Την περιγραφή των θεωρητικών πεδίων της επιστήμης, των τεχνικών και μεθόδων που εμπλέκονται σε ένα τέτοιο έργο λογισμικού
- Η δοκιμή της εφαρμογής των αλγορίθμων σε ροή τυχαίων εικόνων και η παραγωγή της αντίστοιχης ηχητικής επένδυσης θα χρησιμοποιηθεί ως μέθοδος επίδειξης της δυνατότητας να προκύψει ένα αποτέλεσμα αποδεκτής ακουστικής αισθητικής.
- Η δοκιμή της εφαρμογής θα αποδείξει επίσης ότι κάτι τέτοιο είναι εφικτό να εκτελεστεί σε πραγματικό χρόνο και χωρίς αντιληπτή υστέρηση παρουσίασης της ροής ακόμη και όταν ο αριθμός των εικόνων και το μέγεθος τους (υπό την έννοια μεγέθους των αρχείων) είναι σχετικά μεγάλος.
- Δεν υπάρχει κάποιο αντικειμενικό κριτήριο μέτρησης της αποτελεσματικότητας των αλγορίθμων και της εφαρμογής γενικότερα, όμως η απρόσκοπτη λειτουργία καθώς και το επίπεδο αρμονικότητας του αποτελέσματος μπορούν να γίνουν αντιληπτές.

Κατηγορίες Χρηστών στους οποίους απευθύνεται το σύστημα

Κανονικοί χρήστες

Με τον όρο “κανονικό χρήστη” εννοούμε οποιοδήποτε χρήστη θέλει να χρησιμοποιήσει το λογισμικό ως έχει, προκειμένου να παρακολουθήσει μια ροή στατικών εικόνων

αποτελούμενη από εικόνες που έχει επιλέξει ακούγοντας ταυτοχρόνως και την ηχητική επένδυση που έχει δημιουργηθεί χρησιμοποιώντας δεδομένα που εξήχθησαν από τις προβαλλόμενες φωτογραφίες σε πραγματικό χρόνο. Η κατηγορία των κανονικών χρηστών αναμένεται να έχει εύλογες απαιτήσεις όσον αφορά την ταχύτητα απόκρισης του συστήματος, την ευχρηστία και την ευκολία με την οποία θα είναι σε θέση να αλληλεπιδράσει με το σύστημα και επίσης το επίπεδο αρμονίας και αποδεκτής ακουστικής αισθητικής του αποτελέσματος.

Χρήστες ή ερευνητές που ενδιαφέρονται για το πεδίο πειραματισμού όσον αφορά την αυτοματοποιημένη σύνθεση ηχητικού ερεθίσματος με χρήση ως εισόδου δεδομένα εικόνας

Ο σκοπός δημιουργίας του εν λόγω συστήματος λογισμικού δεν είναι πρωτίστως η χρήση του για λόγους πειραματισμού. Για παράδειγμα δεν θα παρέχονται πρωτογενώς στο χρήστη δυνατότητες παραμετροποίησης του ήχου κατά τη διάρκεια της ροής εικόνων. Οι δυνατότητες παραμετροποίησης της ροής περιορίζονται μόνο στο οπτικό τμήμα της ροής και είναι εκείνες που ισχύουν στις περισσότερες εφαρμογές που διατίθενται εμπορικά ή δωρεάν για την προβολή μιας ροής στατικών εικόνων (slideshow) χωρίς ήχο. Όπως για παράδειγμα η ταχύτητα εναλλαγής από τη μια εικόνα στην άλλη ή η επιλογή του χρησιμοποιούμενου εφέ μετάβασης από τη μια εικόνα στην άλλη. Ωστόσο υπάρχει δυνατότητα αξιοποίησης των δεδομένων που εξάγονται από την εικόνα για την παραγωγή ήχου με διαφορετικούς τρόπους μέσω προγραμματισμού (βλ. Κεφάλαιο σχετικό με την εφαρμογή παραγωγής ήχου Csound) και αποτελεί πεδίο πειραματισμού εξαιρετικά μεγάλου εύρους, το οποίο δε μπορούσε να καλυφθεί μέσω της παρούσας διατριβής.

3 ΑΝΑΛΥΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

Εισαγωγή

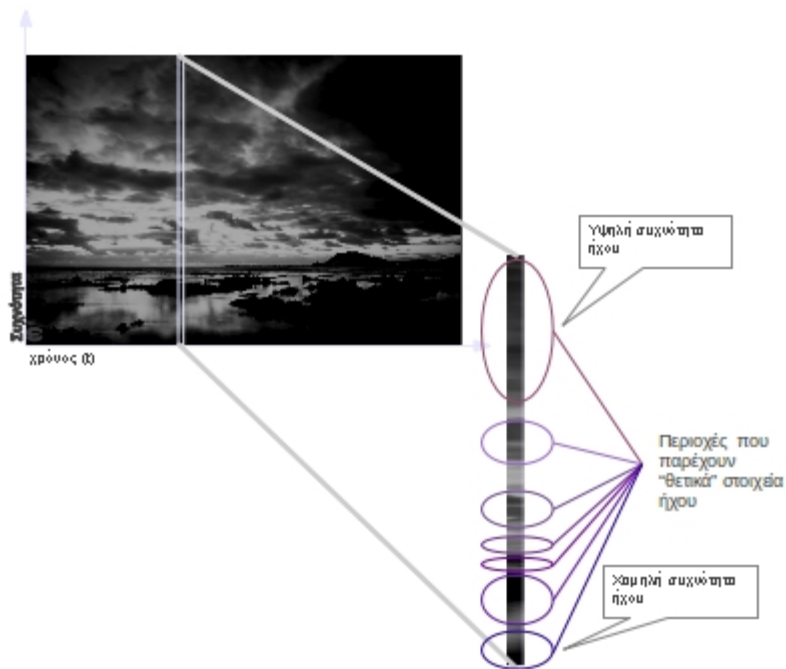
Το πρώτο μέρος της ανάλυσης του συστήματος περιλαμβάνει τα κριτήρια που χρησιμοποιήθηκαν για την επιλογή των αλγορίθμων εξαγωγής δεδομένων από εικόνα. Στη συνέχεια παρουσιάζουμε τις λειτουργικές απαιτήσεις του συστήματος λογισμικού.

Κριτήρια επιλογής αλγορίθμων

Σε εφαρμογές αυτού του τύπου όπου το αποτέλεσμα κρίνεται κατά κύριο λόγο σε υποκειμενικό επίπεδο η επιλογή αλγορίθμων υπαγορεύεται εξίσου από λειτουργικές αλλά και μη λειτουργικές απαιτήσεις. Εξαιτίας του παράγοντα του υποκειμενισμού, δεν υπάρχουν διαθέσιμες στη διεθνή βιβλιογραφία λειτουργικές και δοκιμασμένες προτάσεις πάνω σε τέτοιου είδους αλγόριθμους με επισήμως καταγεγραμμένες επιδόσεις και “αποδεδειγμένο” αισθητικό αποτέλεσμα. Μετά από αρκετά εκτενή έρευνα, βρέθηκε ότι περισσότερες υπάρχουσες εφαρμογές εστιάζουν κυρίως στο τμήμα του επιστημονικού και πειραματικού ενδιαφέροντος που προκύπτει από την αντιστοίχιση δεδομένων εικόνας με δεδομένα ήχου χωρίς να δίνεται κάποια βαρύτητα στην αρμονικότητα του ηχητικού αποτελέσματος. Επομένως η εύρεση κατάλληλων αλγορίθμων που θα μπορούσαν , ξεκινώντας από δεδομένα εικόνας, τελικά να οδηγήσουν σε ένα αποδεκτό (ακουστικό) αισθητικό αποτέλεσμα ήταν μέρος της εργασίας αυτής.

Όπως είδαμε παραπάνω, στις ιατρικές συσκευές που λειτουργούν υποστηρικτικά σε ασθενείς με προβλήματα όρασης, μια προσέγγιση είναι εκείνη που αξιοποιεί τα εικονοστοιχεία ως είσοδο για την παραγωγή ήχου με βάση τις καθ' ύψος συντεταγμένες. Αυτή αντιστοιχίζει τα εικονοστοιχεία που βρίσκονται στις ανώτερες περιοχές τις εικόνας με υψηλές συχνότητες ενώ εκείνα που βρίσκονται στις κατώτερες ζώνες της με τις χαμηλές και -βάσει παραδοχής- η παραγωγή ή η μη-παραγωγή ήχου αποφασίζεται με βάση τον σχετικό χρωματικό τόνο (π.χ. οι σκουρότεροι χρωματικοί τόνοι οδηγούν σε παραγωγή ήχου ενώ οι ανοιχτόχρωμοι όχι). Μία προσαρμογή αυτής της ιδέας σε ροή στατικών εικόνων θα μπορούσε να πραγματοποιηθεί ως εξής:

Η εικόνα σαρώνεται κατά μήκος και δεδομένα εξάγονται για κάθε κάθετη τομή (προφίλ) της εικόνας. Τα “σκουρότερα” εικονοστοιχεία -βάσει παραδοχής- θεωρούνται ως εκείνα που θα αξιοποιηθούν για την παραγωγή του ήχου. Ο παραγόμενος ήχος, δε, είναι σύνθεση των συχνοτήτων που παράγονται αντιστοιχίζοντας τα εικονοστοιχεία στις ανώτερες περιοχές της εικόνας με υψηλές συχνότητες και εκείνα που βρίσκονται στις κατώτερες ζώνες σε χαμηλές συχνότητες. Μια απεικόνιση αυτής της αρχής φαίνεται παρακάτω:



Εικόνα 14 Κάθετες τομές/προφίλ και η αξιοποίησή τους ως δεδομένων ήχου

Εκτός από τα δεδομένα εικόνας τα οποία σχετίζονται με τη γεωμετρία του απεικονιζόμενου θέματος, υπάρχουν και εκείνα που αφορούν στο σύνολο της εικόνας, όπως για παράδειγμα η φωτεινότητα, η αντίθεση και ο κορεσμός τα οποία θα ονομάσουμε στα πλαίσια της παρούσας εργασίας “γενικά”. Στην κατηγορία των γενικών δεδομένων της εικόνας μπορούν να ενταχθούν και άλλα στοιχεία της εικόνας όπως η συνολική σχετική αναλογία σκοτεινών-φωτεινών περιοχών αλλά και τα η αναλογία κόκκινου-πράσινου-μπλε.

Ο περιορισμός των “προς ηχητική αξιοποίηση” δεδομένων εικόνας στα γενικά στοιχεία της εικόνας δε φαίνεται να μπορεί να μας δώσει ποικιλομορφία στα παραγόμενα δεδομένα ήχου μιας και αντιστοιχεί μια μόνο αριθμητική τιμή ανά εικόνα για αυτά τα δεδομένα.

Η προσπάθεια ένταξης στα ηχητικά δεδομένα και επιμέρους στοιχείων του απεικονιζόμενου θέματος, όπως εκείνων που προκύπτουν αξιοποιώντας τις κάθετες τομές της, που είδαμε πιο πριν, μας παρέχει τη δυνατότητα εξαγωγής περισσότερο δυναμικών και ποικιλόμορφων δεδομένων. Ωστόσο η προσπάθεια αντιστοίχισης των σκούρων

γεωμετρικών στοιχείων μιας εικόνας με ήχους διαφόρων συχνοτήτων με βάση τη θέση τους στην εικόνα μπορεί να οδηγήσει στην παραγωγή θορύβου μιας και στις μεγάλες ηχητικά αξιοποιήσιμες, στην περίπτωση της παραδοχής του παραδείγματος σκουρόχρωμες, περιοχές/προφίλ μπορεί να έχουμε παραγωγή συχνότητας με πολλά μη-αρμονικά συνθετικά.

Επομένως διαφαίνεται αρκετά νωρίς ότι η αξιοποίηση και των επιμέρους στοιχείων της εικόνας μπορεί να οδηγήσει σε ένα πιο ολοκληρωμένο ηχητικό αποτέλεσμα. Όμως χρειαζόμαστε μια μέθοδο για να αποκτήσουμε μεγαλύτερο έλεγχο στον αριθμό των αρμονικών που μπορούν να προκύψουν από το κάθε προφίλ. Ένα άλλο πρόβλημα προκύπτει από το γεγονός ότι η ποσότητα των δεδομένων που μπορεί να προκύψει από μια εικόνα είναι πολύ υψηλή. Για μια εικόνα με ανάλυση 600x400 (πλάτος x ύψος) το κάθε προφίλ/κάθετη τομή εικόνας θα περιλαμβάνει δεδομένα για 400 εικονοστοιχεία και -εάν αξιοποιηθεί όλο το μήκος της εικόνας- συνολικά θα πρέπει να εξαχθούν 600 τέτοια προφίλ. Εάν κατά την προβολή της ροής εικόνων, η κάθε εικόνα παραμένει στην οθόνη για 5 δευτερόλεπτα, τότε θα αντιστοιχούν 120 προφίλ εικόνας/δευτερόλεπτο. Η “πυκνότητα” αυτή ροής δεδομένων οδηγεί σε αρκετά δύσκολη διαχείρισή τους υπό τη μορφή δεδομένων ήχου. Το πρόβλημα αυτό, στα πλαίσια της παρούσας διπλωματικής εργασίας, αντιμετωπίστηκε με την εφαρμογή ενός κλασσικού πλέον αλγορίθμου, του αλγορίθμου ανίχνευσης ακμών του John Canny [26] ο οποίος υλοποιείται σε ένα σύνολο διαδοχικών βημάτων.

Αλγόριθμος ανίχνευσης ακμών Canny

Ο σκοπός των αλγορίθμων ανίχνευσης ακμών, γενικά, είναι να ελαττωθεί η ποσότητα της πληροφορίας σε μια εικόνα διατηρώντας ωστόσο τα δομικά στοιχεία της εικόνας ώστε να μπορούν να χρησιμοποιηθούν για περαιτέρω επεξεργασία. Ο αλγόριθμος ανίχνευσης ακμών του Canny πρόκειται για έναν από τους πρώτους (1986) αλγορίθμους για τη συγκεκριμένη εργασία και τα αποτελέσματά του έχουν επαληθευτεί εκτενώς και εξακολουθεί να χρησιμοποιείται στον τομέα της έρευνας. Στόχος του Canny ήταν να αναπτύξει έναν αλγόριθμο ο οποίος να παρουσιάζει καλές επιδόσεις όσον αφορά τα παρακάτω κριτήρια:

- Ανίχνευση: Η πιθανότητα ανίχνευσης πραγματικών ακμών πρέπει να μεγιστοποιείται ενώ η πιθανότητα λανθασμένης ανίχνευσης ως ακμών πρέπει να

ελαχιστοποιείται. Αυτό αντιστοιχεί σε μεγιστοποίηση του λόγου σήμα-θόρυβος.

- Τοπικότητα: Οι ακμές που ανιχνεύονται πρέπει να βρίσκονται όσο πλησιέστερα γίνεται στις πραγματικές ακμές.
- Μοναδικότητα: μια πραγματική ακμή δεν πρέπει να ανιχνεύεται ως ακμή περισσότερες από μια φορές.

Ο αλγόριθμος υλοποιείται σε 5 διαδοχικά στάδια:

1. Αλλαγή των χρωμάτων της εικόνας σε διαβαθμίσεις του γκριζου (grayscale) και φιλτράρισμα της ώστε να απομακρυνθεί ο οπτικός θόρυβος
2. Υπολογισμός της κλίσης (ανάδελτα \square , gradient) έντασης με χρήση μασκών Sobel. Επισημαίνονται οι ακμές, δηλαδή τα σημεία όπου η κλίση έντασης παρουσιάζει μέγιστα.
3. Καταστολή μη-μεγίστων τιμών. Μόνο τα τοπικά μέγιστα πρέπει να σημειωθούν ως ακμές.
4. Κατωφλίωση (threshold) Υστέρησης. Πιθανές ακμές καθορίζονται μέσω κατωφλίωσης.
5. Ανίχνευση ακμών με υστέρηση. Οι τελικές ακμές καθορίζονται καταστέλλοντας όσες ακμές δεν είναι συνδεδεμένες με κάποια “ισχυρή” ακμή

Αναλυτικά η διαδικασία αναφέρεται παρακάτω:

1.Αλλαγή των χρωμάτων της εικόνας σε διαβαθμίσεις του γκριζου - φιλτράρισμα

Η εικόνα μετατρέπεται από έγχρωμη σε ασπρόμαυρη για να απλοποιηθούν οι υπολογισμοί κατά τα υπόλοιπα στάδια. Στα μεταγενέστερα στάδια, αντί να λαμβάνονται υπόψη και οι τρεις “συνιστώσες” χρώματος RGB, αρκεί ο έλεγχος μιας μόνον από αυτές. Κατόπιν, η εφαρμογή φίλτρου στην προς επεξεργασία εικόνα είναι απαραίτητη, ειδικά στις φωτογραφίες, γιατί συχνά έχουν κάποια ποσότητα θορύβου. Για να αποφύγουμε την ανίχνευση του θορύβου ως ακμές απομακρύνουμε το θόρυβο αυτό προτού να συνεχίσουμε με τα υπόλοιπα βήματα. Το φίλτρο που εφαρμόζεται είναι το Gaussian. Το

κέλυφος του Gaussian φίλτρου με απόκλιση $\sigma = 1.4$ φαίνεται στη σχέση (1):

$$B = \frac{1}{159} \cdot \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} \quad (1)$$

2. Εύρεση κλίσης έντασης

Ο αλγόριθμος του Canny ανιχνεύει ακμές στα σημεία όπου η μεταβολή της έντασης του χρώματος γίνεται μέγιστη. Οι περιοχές μέγιστης μεταβολής προσδιορίζονται από την “κλίση” της εικόνας. Η κλίση (\square) για κάθε εικονοστοιχείο προσδιορίζεται με την εφαρμογή των τελεστών Sobel. Αυτό το βήμα πραγματοποιείται σε δύο φάσεις. Κατά την πρώτη φάση βρίσκουμε την κλίση στους άξονες των x- και y- εφαρμόζοντας τα κελύφη (2):

$$K_{GX} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$
$$K_{GY} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (2)$$

Στη συνέχεια η βάθμωση κλίσης (ένταση ακμής) μπορεί να προσδιοριστεί ως μια ευκλείδειος απόσταση η οποία μπορεί να μετρηθεί με τη βοήθεια του πυθαγορείου θεωρήματος (3) ή στην απλοποιημένη μορφή (4):

$$|G| = \sqrt{G_x^2 + G_y^2} \quad (3)$$

$$|G| = |G_x| + |G_y| \quad (4)$$

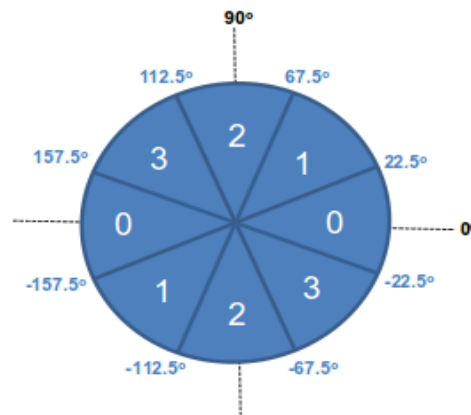
Όπου τα G_x και G_y είναι αντιστοίχως οι βαθμώσεις κατά τον x- και y-. Τελικά η διεύθυνση των ακμών είναι αυτή που αποθηκεύεται (για να χρησιμοποιηθεί αργότερα) με τη βοήθεια της σχέσης (5) και που τελικά αντιπροσωπεύει την εικόνα της βάθμωσης κλίσης:

$$\theta = \arctan\left(\frac{|G_y|}{|G_x|}\right) \quad (5)$$

3. Καταστολή μη-μεγίστων τιμών

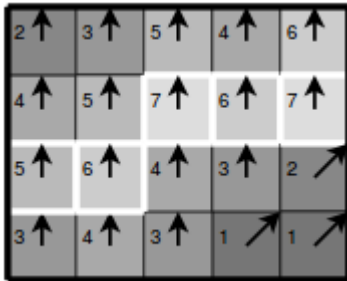
Σε αυτό το βήμα ο σκοπός είναι να μετατρέψουμε τις “φιλτραρισμένες” ακμές στην εικόνα βάθμωσης κλίσης σε “ισχυρές” ακμές. Αυτό το επιτυγχάνουμε διατηρώντας όλα τα τοπικά μέγιστα στην εικόνα βάθμωσης κλίσης και αφαιρώντας όλα τα υπόλοιπα. Ο αλγόριθμος εφαρμόζεται σε κάθε σημείο της εικόνας βάθμωσης κλίσης:

- 1) Στρογγυλοποιούμε τη βάθμωση της διεύθυνσης θ στην πλησιέστερη γωνία 45° ώστε η διεύθυνση όλων των γειτονικών στοιχείων (κάθε στοιχείο περιβάλλεται από 8 γειτονικά) να είναι τελικά βαθμωτή σε βήματα των 45° (Εικόνα 15).



- 2) Συγκρίνουμε την ένταση της ακμής του εικονοστοιχείου με τη δύναμη ακμής των εικονοστοιχείων με τα οποία συνορεύει υπό αυτή τη γωνία. Για παράδειγμα, αν η γωνία θ είναι 90° τότε εξετάζουμε το “βόρειο” και “νότιο” γειτονικό εικονοστοιχείο.
- 3) Εάν η ένταση του υπο εξέταση εικονοστοιχείου είναι μεγαλύτερη, τότε διατηρείται. Στην αντίθετη περίπτωση αφαιρείται.

Παράδειγμα εφαρμογής της καταστολής μη-μέγιστων τιμών φαίνεται στην Εικόνα 16. Τα περισσότερα εικονοστοιχεία έχουν βάθμωση κλίσης με διεύθυνσης βόρεια. Συνεπώς η σύγκριση της έντασης για αυτά πραγματοποιείται σε σχέση με τα εικονοστοιχεία που βρίσκονται πάνω και κάτω από αυτά. Τα εικονοστοιχεία που βρέθηκαν να έχουν τις μέγιστες τιμές έχουν σημειωθεί με λευκό περίγραμμα.



Εικόνα 16 Η ένταση των ακμών απεικονίζεται χρωματικά και αριθμητικά, ενώ η κατεύθυνση της βάθμωσης δείχνεται με βέλη. Τα εικονοστοιχεία που ταυτοποιούνται ως ακμές έχουν σημειωθεί με λευκό περίγραμμα

4. Κατωφλίωση υστέρησης

Τα εικονοστοιχεία-ακμές που έχουν απομείνει μετά το προηγούμενο βήμα είναι ακόμη σημειωμένα με την έντασή τους ένα-προς-ένα. Πολλά από αυτά είναι πραγματικές ακμές. Όμως κάποια μπορεί απλώς να είναι θόρυβος ή απλώς έντονη διακύμανση χρώματος. Αυτό το βήμα του αλγορίθμου αναλαμβάνει να διακρίνει μεταξύ των δύο. Κατά το στάδιο αυτό επιτρέπεται η διατήρηση μόνο όσων ακμών είναι πάνω από το κατώφλιο έντασης. Πιο συγκεκριμένα εφαρμόζεται διπλή κατωφλίωση: τα εικονοστοιχεία-ακμές

των οποίων η έντασή τους είναι υψηλότερη από το άνω κατώφλιο σημειώνονται ως “ισχυρά”, τα εικονοστοιχεία-ακμές των οποίων η έντασή τους είναι χαμηλότερη από το κάτω κατώφλιο καταστέλλονται ενώ όλα τα ενδιάμεσα σημειώνονται ως “ασθενή”.

5. Ανίχνευση ακμών με υστέρηση

Οι ισχυρές ακμές κατοχυρώνονται και περιλαμβάνονται στην τελική εικόνα. Ασθενείς ακμές περιλαμβάνονται στην τελική εικόνα μόνο εάν είναι συνδεδεμένες με ισχυρές ακμές. Σε αυτό το βήμα, σκοπός είναι η εξαίρεση των ασθενών ακμών εκείνων που προκύπτουν από διακύμανση του χρώματος. Γίνεται δηλαδή η παραδοχή ότι οι ασθενείς ακμές που οφείλονται σε διακύμανση χρώματος θα έχουν τυχαία κατανομή στην επιφάνεια της εικόνας και συνεπώς δεν θα συνδέονται με ισχυρές ακμές.

Παρακάτω φαίνεται ένα παράδειγμα της επίδρασης του αλγόριθμου:



Εικόνα 17 Παράδειγμα επίδρασης του αλγορίθμου ανίχνευσης ακμών Canny

Η πολυπλοκότητα εκτέλεσης για το κάθε στάδιο υπολογίζεται ως εξής (στους παρακάτω υπολογισμούς το n συμβολίζει τον αριθμό των εικονοστοιχείων):

- Η αντικατάσταση των χρωμάτων της εικόνας από διαβαθμίσεις του γκριζου “κοστίζει” n επαναλήψεις, ενώ η εφαρμογή του Gaussian φίλτρου εκτελείται σε $n + 255 + n$. Επομένως η πολυπλοκότητα για αυτό το στάδιο είναι $\Theta(n)$.

- Η εύρεση της κλίσης έντασης έχει πάλι γραμμική πολυπλοκότητα γιατί ο μέγιστος αριθμός επαναλήψεων είναι $3 \times (2xn) + n$ μιας και χρησιμοποιούμε δύο τελεστές/μάσκες Sobel. Οπότε και αυτό το στάδιο έχει γραμμική πολυπλοκότητα $\Theta(n)$.
- Η καταστολή μη-μεγίστων τιμών απαιτεί συνολικά $n + n$ επαναλήψεις οπότε και εδώ έχουμε γραμμική πολυπλοκότητα $\Theta(n)$.
- Η εφαρμογή της υστέρησης κοστίζει στη χειρίστη περίπτωση n επαναλήψεις συνεπώς εξάγεται ότι για αυτό το στάδιο η πολυπλοκότητα είναι $\Theta(n)$.
- Η ανίχνευση ακμών με υστέρηση εκτελείται συνολικά σε n επαναλήψεις οπότε και σε αυτό το τελευταίο στάδιο η πολυπλοκότητα είναι επίσης γραμμική και ίση με n .

Τελικά βλέπουμε ότι οι χρόνοι εκτέλεσης είναι συνολικά ευθέως ανάλογοι του μεγέθους της εικόνας προς επεξεργασία.

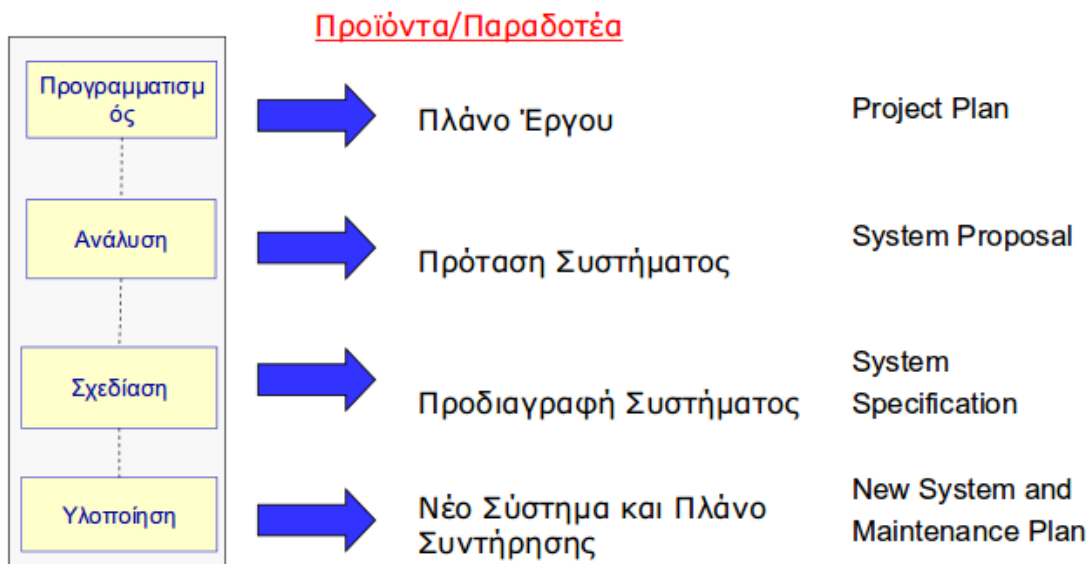
4 ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΑΠΤΥΞΗΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

Εισαγωγή

Σε αυτή την ενότητα θα αναφέρουμε τη μεθοδολογία η οποία ακολουθήθηκε για να αναπτυχθεί το λογισμικό και περιγράφονται τα αναμενόμενα αποτελέσματα από την ανάπτυξή του. Θα παραθέσουμε τους λόγους για τους οποίους επιλέχθηκε η συγκεκριμένη μεθοδολογία ανάπτυξης. Επίσης θα αναφερθούμε στα εργαλεία και το περιβάλλον ανάπτυξης που χρησιμοποιήθηκαν αλλά και τους λόγους για τους οποίους τα συγκεκριμένα επιλέχθηκαν. Στη συνέχεια θα κάνουμε εμφανείς σε υψηλό επίπεδο θεώρησης την αρχή λειτουργικότητας του συστήματος λογισμικού που αναπτύχθηκε. Θα αναφερθούν επίσης οι λειτουργικές απαιτήσεις αλλά και οι περιπτώσεις χρήσης. Θα γίνει τέλος αναφορά στις μη-λειτουργικές απαιτήσεις και μια σύντομη συζήτηση σχετικά με τους κανόνες που πρέπει να εφαρμόζονται σε εφαρμογές αυτού του τύπου ώστε να εξασφαλίζεται η απρόσκοπτη αλληλεπίδραση με το χρήστη.

Μεθοδολογίες ανάπτυξης λογισμικού

Όλες οι μεθοδολογίες ανάπτυξης λογισμικού έχουν τον ίδιο τελικό σκοπό: τη δημιουργία λογισμικού το οποίο να ανταποκρίνεται στις απαιτήσεις και τις ανάγκες του χρήστη με το μέγιστο βαθμό ποιότητας και με το ελάχιστο κόστος. Επίσης, σε όλες οι μεθοδολογίες μπορούμε να εντοπίσουμε κάποιες θεμελιώδεις φάσεις: προγραμματισμός, ανάλυση, σχεδίαση, υλοποίηση. Κάθε φάση αποτελείται από υπο-φάσεις ή βήματα και σε κάθε ένα από αυτά τα βήματα υπάρχει ένας αριθμός από παραδοτέα. Παρακάτω φαίνονται τα παραδοτέα της κάθε φάσης:



Ωστόσο, οι μεθοδολογίες διαφέρουν όσον αφορά τα βήματα που αποτελούν την κάθε φάση ή ακόμη και τη σειρά με την οποία πρέπει να εκτελεστούν. Θα μπορούσαμε να κατηγοριοποιήσουμε τις μεθοδολογίες ανάπτυξης ως εξής [27]:

Απουσία μεθοδολογίας (build-and-fix ή ad-hoc)

Κατά την προσέγγιση αυτή δεν ακολουθείται κάποια συγκεκριμένη διαδικασία για την ανάπτυξη του λογισμικού. Οι απαιτήσεις μετατρέπονται άμεσα σε κώδικα χωρίς πρώτα να ερμηνευτούν ως προδιαγραφές. Ο οποίος συνήθως ελέγχεται απευθείας από τον πελάτη και τυχόν προβλήματα επιλύονται από την ομάδα ανάπτυξης προτού το προϊόν παραδοθεί εκ νέου στον πελάτη. Ο κύκλος αυτός επαναλαμβάνεται έως ότου

ικανοποιηθεί ο πελάτης ενώ όταν ολοκληρωθεί ο κύκλος ζωής του λογισμικού αυτό αποσύρεται.

Δομημένη Σχεδίαση (structured design)

Καταρράκτης (waterfall), παράλληλη (parallel)

Σύμφωνα με τις μεθοδολογίες αυτές δίνεται έμφαση στο σχεδιασμό πριν από κάθε φάση και συνίσταται στην συνεχή επανάληψη των βημάτων: ανάλυση απαιτήσεων (requirements)- επαλήθευση, καθορισμός προδιαγραφών (specifications) - επαλήθευση, σχεδιασμός (construction) - επαλήθευση, υλοποίηση (implementation) - έλεγχος, συνένωση κώδικα (έλεγχος), εφαρμογή (deployment). Καμία φάση δεν θεωρείται ολοκληρωμένη πριν γραφτεί και εγκριθεί η τεκμηρίωσή της. Γνώρισε μεγάλη απήχηση κατά την ανάπτυξη των πρώτων εμπορικών συστημάτων λογισμικού ενώ πριν να εφαρμοστεί για έργα λογισμικού εφαρμοζόταν σε έργα υλισμικού. Λόγω των σαφώς καθορισμένων και αυστηρών σημείων ελέγχου, δεν ανταποκρίνεται πολύ αποτελεσματικά στις αλλαγές απαιτήσεων από την πλευρά του πελάτη. Η αδυναμία της άρχισε να γίνεται έντονα ορατή όταν τα συστήματα λογισμικού αυξήθηκαν σε μέγεθος.

Εξελικτικές/Ταχείας ανάπτυξης εφαρμογής (evolutionary / rapid application development (RAD))

Στις μεθοδολογίες τύπου RAD δίνεται έμφαση στην γρήγορη πρωτοτυποποίηση ενώ ο προγραμματισμός (planning) παραμένει σε μινιμαλιστικά επίπεδα. Η φάση του προγραμματισμού σχεδόν αλληλεπικαλύπτεται με τη φάση της συγγραφής του κώδικα. Αυτή η μέθοδος επιτρέπει γρήγορη συγγραφή κώδικα και διαχειρίζεται καλά τις αλλαγές στις λειτουργικές απαιτήσεις. Αντενδείκνυται όμως στις περιπτώσεις όπου το προϊόν είναι υψηλού τεχνικού κινδύνου και πολυπλοκότητας.

Ευέλικτη Ανάπτυξη (Agile Development)

Στα πλαίσια αυτής της μεθοδολογίας η ανάπτυξη πραγματοποιείται σε μικρά επαναλαμβανόμενα τμήματα και με αθροιστικό τρόπο. Δίδεται ιδιαίτερη έμφαση στην αλληλεπίδραση μεταξύ των μελών μιας ομάδας ανάπτυξης αλλά και μεταξύ των διαφορετικών ομάδων ανάπτυξης. Η αλληλεπίδραση αξιολογείται υψηλότερα από τις

διαδικασίες και τα εργαλεία ανάπτυξης και δεν δίδεται ιδιαίτερη έμφαση στην τεκμηρίωση του κώδικα αλλά στην ορθή λειτουργία του. Οι αλλαγές στις απαιτήσεις είναι ευπρόσδεκτες ακόμη και κατά τη διάρκεια των τελευταίων σταδίων ανάπτυξης. Αυτό είναι εφικτό εν μέρει και λόγω του ότι κατά την ευέλικτη ανάπτυξη ο καταμερισμός του χρόνου ανάπτυξης γίνεται με μικρές χρονικές καταταμήσεις (διάρκειας π.χ. δύο εβδομάδων) τα οποία ονομάζονται sprints και στο τέλος των οποίων παραδίδεται ένα μικρό τμήμα λειτουργικού κώδικα. Διαφοροποιήθηκε από τις υπάρχουσες μεθοδολογίες όσον αφορά την εστίαση της συνεχούς παραγωγής κώδικα [28]. Σε αυτή τη κατηγορία μεθοδολογιών εντάσσεται και το eXtreme Programming (XP) το οποίο διέπεται από τις αξίες:

- Επικοινωνία
- Απλότητα
- Ανατροφοδότηση (feedback)
- Θάρρος (courage)

και που δίνει ιδιαίτερη βαρύτητα στην συνεχή και στενή επαφή με τους τελικούς χρήστες του λογισμικού, τη συνεργασία ανά δύο (pairing) κατά την ανάπτυξη του λογισμικού αλλά και στις συνεχείς δοκιμές (testing) του κώδικα

Άλλη μεθοδολογία που ανήκει σε αυτήν την κατηγορία είναι και εκείνη της απέριτης ανάπτυξης (Lean Software Development). Η μεθοδολογία, μεταξύ άλλων, εστιάζει στα παρακάτω:

- Κατάργηση των μη παραγωγικών διαδικασιών και μεθόδων
- Ενίσχυση του τεχνικού υπόβαθρου μέσω της εκμάθησης τεχνικών και εργαλείων ανάπτυξης μελετώντας τις τεχνολογίες που σχετίζονται με αυτό
- Μετατόπιση των κρίσιμων δεσμεύσεων προς τον πελάτη σε μεταγενέστερο χρονικό σημείο
- Διάθεση παραδοτέων όσο το δυνατόν πιο γρήγορα
- Συνειδητοποίηση του συνόλου του περιβάλλοντος μέσα στο οποίο καλείται να λειτουργήσει το λογισμικό

Σε πραγματικές συνθήκες (στη βιομηχανία ανάπτυξης λογισμικού), οι μεθοδολογίες και

διαδικασίες ανάπτυξης, συνήθως υπαγορεύονται από τον οργανισμό ανάπτυξης του λογισμικού και δεν αποτελούν προσωπική επιλογή του αναλυτή/προγραμματιστή. Η απόφαση για την υιοθέτηση κάποιας μεθοδολογίας ανάπτυξης λαμβάνεται με βάση μια πληθώρα παραμέτρων. Μερικές από αυτές είναι οι παρακάτω:

- Μέγεθος του έργου/έργων

Συχνά είναι μη-αποδοτικό το να επιλεγεί κάποια “βαριά” μεθοδολογία για μικρά έργα.

- Μέγεθος των ομάδων ανάπτυξης

Μικρές ομάδες ανάπτυξης συνήθως είναι πιο αποδοτικές όταν γίνεται χρήση της Agile μεθοδολογίας.

- Βαθμός πολυπλοκότητας του έργου

Είδαμε, για παράδειγμα, ότι η μεθοδολογία του “Καταρράκτη” δεν διαχειρίζεται πάντοτε με αποτελεσματικό τρόπο έργα με υψηλό βαθμό πολυπλοκότητας

- Βαθμός δραστηριοποίησης του οργανισμού στις διαφορετικές φάσεις του κύκλου ανάπτυξης του προϊόντος

Μπορεί, λόγω χάριν, ο οργανισμός να εμπλέκεται μόνο στον έλεγχο και επαλήθευση του προϊόντος.

- Γεωγραφική τοποθέτηση των ομάδων ανάπτυξης

Ομάδες που βρίσκονται στον ίδιο φυσικό χώρο έχουν περισσότερες πιθανότητες να αποδώσουν εργαζόμενες με μεθοδολογίες που προάγουν την προσωπική επαφή (Agile) ενώ οργανισμοί ανάπτυξης λογισμικού με απομεμακρυσμένες ομάδες ενδεχομένως είναι πιο αποτελεσματικές με μεθοδολογίες που ελαχιστοποιούν την ανάγκη για αλληλεπίδραση (Καταρράκτης)

- Εξωγενείς παράγοντες (λ.χ. όταν η μεθοδολογία ανάπτυξης αποτελεί σημείο ελέγχου για την πιστοποίηση από κάποιον εξωτερικό οργανισμό διασφάλισης ποιότητας)

Συχνά δε, ο οργανισμός/επιχείρηση θέτει και περιορισμούς όσον αφορά το περιβάλλον και τα εργαλεία ανάπτυξης ή ακόμα και τα καθορίζει με συγκεκριμένο τρόπο.

Επιλογή μεθοδολογίας για την ανάπτυξη του συστήματος

Στα πλαίσια της παρούσας εργασίας, είχαμε την “πολυτέλεια” να επιλέξουμε τη

μεθοδολογία που θα χρησιμοποιηθεί αλλά και το περιβάλλον και τα εργαλεία ανάπτυξης. Η μεθοδολογία που επιλέχθηκε είναι εκείνη της απεριττής ανάπτυξης. Οι λόγοι για αυτήν την απόφαση είναι οι παρακάτω [29]:

-Απαιτείται εκμάθηση του περιβάλλοντος, των τεχνικών και των εργαλείων ανάπτυξης σε ικανοποιητικό βαθμό

Αυτή η αρχή/προϋπόθεση είναι μάλλον αυτονόητη σε ένα έργο το οποίο αποτελεί μέρος κάποιας εκπαιδευτικής διαδικασίας. Στη δική μας περίπτωση, η ανάγκη για εξοικείωση με αυτό τον χώρο τεχνογνωσίας στάθηκε ακόμη πιο επιτακτική μιας και δεν υπήρχε καμία πρότερη επαφή με κώδικα και τεχνικές ανάπτυξης συστήματος λογισμικού σχετικού με προβολή, διαχείριση και επεξεργασία εικόνας αλλά ούτε με συστήματα επεξεργασίας και αναπαραγωγής ήχου. Η έλλειψη υποδομής σε αυτές τις περιοχές από την πλευρά του αναλυτή/προγραμματιστή παρουσιάζει κάποια αναλογία με την περίπτωση οργανισμού ο οποίος εμπλέκεται στην ανάπτυξη λογισμικού σε άγνωστη τεχνική περιοχή. Σε τέτοιου είδους έργα η εστίαση στο γνωστικό τμήμα της ανάπτυξης ενισχύεται μέσω της εφαρμογής του lean software development.

-Η ανάπτυξη εξ' αρχής ενός συστήματος το οποίο θα αναλάμβανε τόσο τη διαχείριση του σκέλους “διαχείριση εικόνας, προβολή στατικής ροής εικόνων, εξαγωγή δεδομένων εικόνας αξιοποιήσιμων για την παραγωγή ηχητικής επένδυσης” όσο και του σκέλους “αξιοποίηση των δεδομένων εικόνας για την παραγωγή ήχου” φάνηκε από νωρίς ότι απαιτεί ανθρωποέργο μεγαλύτερο από εκείνο που μπορούσε να καταναλωθεί στα χρονικά πλαίσια της εργασίας. Για αυτό το λόγο προτιμήθηκε η οδός της επαναχρησιμοποίησης κώδικα, στρατηγικής που είναι σύμφωνη με τη φιλοσοφία του Agile

-Η υλοποίηση μπορεί να κατακερματιστεί σε μικρότερα τμήματα λειτουργικότητας τα οποία μπορούν να αναπτυχθούν σε σύντομο χρονικό διάστημα και να ελεγχθούν/επαληθευτούν μεμονωμένα.

Αυτή η αρχή βοήθησε στην επαλήθευση των βασικών αρχών της σύλληψης σε σχετικά πρώιμο στάδιο ώστε να επαληθευτούν σε αρχικό στάδιο και η ιδέα, οι βασικές γραμμές αρχιτεκτονικής σχεδίασης του συστήματος αποφεύγοντας να καταλήξουμε ενδεχομένως σε σχεδιαστικό αδιέξοδο σε μεταγενέστερα στάδια ανάπτυξης.

-Προσήλωση σε λειτουργικό λογισμικό και όχι σε εκτενή τεκμηρίωση.

Το υπό ανάπτυξη σύστημα προσθέτει μιαν επιπλέον διάσταση σε περιπτώσεις χρήσης με τις οποίες οι περισσότεροι χρήστες είναι αρκετά εξοικειωμένοι (προβολή ροής εικόνων από συλλογή εικόνων του χρήστη). Ως εκ τούτου η εκτενή τεκμηρίωση δεν στάθηκε αυτοσκοπός για την εργασία αυτή. Αντί αυτού πραγματοποιήθηκε εντατική επαλήθευση και έλεγχος του λογισμικού (trial and error) ώστε να εξασφαλιστεί η ποιότητα του.

-Αυτο-οργανούμενη ομάδα

Η μέθοδος της απέριτης ανάπτυξης εστιάζει στην αυτο-οργάνωση της ομάδας. Συνεπώς ευνοεί την ανάπτυξη από έναν μόνο αναλυτή/προγραμματιστή.

5 ΣΧΕΔΙΑΣΗ ΣΥΣΤΗΜΑΤΟΣ

Επιλέχθηκε το όνομα Sound Track Generator (Γεννήτρια Ηχητικού Ίχνους) για το σύστημα λογισμικού που κατασκευάστηκε στα πλαίσια της εργασίας αυτής. Θα χρησιμοποιούμε επίσης τη συντετμημένη εκδοχή του ονόματος STGen για λόγους συνομιίας. Κατατάσσεται στις εφαρμογές ψυχαγωγίας/διασκέδασης. Η λειτουργία της συνοπτικά μπορεί να περιγραφεί ως μια εξελιγμένη προσέγγιση της προβολής ροής στατικών εικόνων (slideshow) κατά την οποία εκτός από το οπτικό ερέθισμα, η εφαρμογή προσφέρει και ηχητικό ερέθισμα παράγοντας με αυτοματοποιημένο τρόπο και αξιοποιώντας δεδομένα που έχουν εξαχθεί από τις προβαλλόμενες εικόνες ηχητική επένδυση για το slideshow.

Θα ξεκινήσουμε από μια συνοπτική περιγραφή των περιπτώσεων χρήσης (Use Cases) που προκύπτουν για μια εφαρμογή αυτού του τύπου και στη συνέχεια θα επεκταθούμε στη περιγραφή της διαδικασίας υλοποίησης.

Οι περισσότεροι χρήστες είναι εξοικειωμένοι με τις ροές στατικών εικόνων (όχι απαραίτητως φωτογραφικών) είτε από προσωπική χρήση κάποιας σχετικής εφαρμογής τοπικά στον υπολογιστή (π.χ. Microsoft Powerpoint, Microsoft Photo Manager, Oracle Openoffice/Libreoffice Impress), η μέσω δικτυακών υλοποιήσεων (Picasa, Flickr, Facebook) είτε ως θεατές σε κάποια διάλεξη, σεμινάριο ή παρουσίαση στην οποία χρησιμοποιήθηκε κάποια σχετική εφαρμογή ως βοήθημα παρουσίασης του υλικού.

Εστιάζοντας αμιγώς στις εφαρμογές οι οποίες προορίζονται για χρήση με φωτογραφικό υλικό, οι απαιτήσεις υψηλού επιπέδου είναι οι ακόλουθες (δεν έχει συμπεριληφθεί ακόμη η επιλογή για προβολή με αυτόματη ηχητική επένδυση):

Ο χρήστης εισάγει ένα αρχείο φωτογραφίας ή ένα φάκελο ο οποίος περιέχει φωτογραφίες. Στη συνέχεια μπορεί να ενεργοποιήσει την προβολή ροής εικόνων και οι φωτογραφίες του να προβάλλονται διαδοχικά στην οθόνη. Ο χρήστης έχει τη δυνατότητα να διακόψει και να επανεκκινήσει τη στατική ροή. Ο χρήστης μπορεί να προσθέσει ή να αφαιρέσει φωτογραφίες από τη ροή εικόνων και να επιλέξει την φωτογραφία από την οποία θα ξεκινήσει η προβολή της ροής εικόνων.

Στα πλαίσια της ανάπτυξης με εφαρμογής απέριπτης μεθοδολογίας δεν χρειάζεται να γίνει εξουθενωτική εξάντληση των πιθανών σεναρίων χρήσης. Ωστόσο, θα χρησιμοποιήσουμε την παραπάνω περιγραφή υψηλού επιπέδου για να περιγράψουμε πιο αναλυτικά τις περιπτώσεις χρήσης λαμβάνοντας αυτή τη φορά υπ' όψιν μας και τη δυνατότητα να αναπαράγεται η ροή και από την ηχητική επένδυση χωρίς όμως να μπούμε σε λεπτομέρειες.

Ο χρήστης μπορεί να επιλέξει είτε με τη βοήθεια συσκευής κατάδειξης είτε μέσω του πληκτρολογίου ένα ή περισσότερα αρχεία ή φακέλους και να τα εισάγει σέρνοντας την προεπιλεγμένη ομάδα εντός του παραθύρου της εφαρμογής. Μπορεί εναλλακτικά να εισάγει τις φωτογραφίες επιλέγοντας απευθείας τις φωτογραφίες και στη συνέχεια επιλέγοντας να τις ανοίξει με την εφαρμογή. Τέλος θα είναι εφικτό ο χρήστης να εισάγει μεμονωμένα φωτογραφίες ή συλλογές φωτογραφιών μέσα από την εφαρμογή επιλέγοντας το πλήκτρο “Open File” ή και να εισάγει φωτογραφίες ή/και συλλογές φωτογραφιών από τον ιστό. Οι εισηγμένες φωτογραφίες θα πρέπει να αποθηκεύονται τοπικά σε βάση δεδομένων για μελλοντική χρήση τους και ο χρήστης θα μπορεί είτε να επιλέξει την τοποθεσία αποθήκευσης τους χειροκίνητα είτε να χρησιμοποιήσει την προεπιλεγμένη τοποθεσία.

Ο χρήστης μπορεί να διαγράψει μια ή περισσότερες φωτογραφίες από τη βάση δεδομένων.

Οι υποστηριζόμενοι τύποι αρχείου πρέπει να καλύπτουν τους πιο διαδεδομένους (raw, png, tiff, jpg).

Ο χρήστης μπορεί να δημιουργήσει μια ή περισσότερες συλλογές από φωτογραφίες και να τις ονοματίσει όπως επιθυμεί.

Ο χρήστης μπορεί να επιλέξει τη δημιουργία ροής εικόνων για οποιαδήποτε από τις συλλογές του. Επίσης μπορεί να επιλέξει τη δημιουργία ροής στατικών εικόνων με αυτόματη δημιουργία ηχητικής επένδυσης για οποιαδήποτε από τις συλλογές του. Η ηχητική επένδυση θα πρέπει να μην προκαλεί έντονη υστέρηση στην εκκίνηση της παρουσίασης της ροής στατικών εικόνων.

Ο χρήστης μπορεί να επιλέξει το ρυθμό με τον οποίο θα προβάλλεται η ροή στατικών εικόνων. Επίσης θα μπορεί να επιλέξει τον τρόπο μετάβασης από τη μια εικόνα στην άλλη κατά την προβολή τους και θα είναι σε θέση να επιλέξει να διακόψει τη ροή ανά πάσα στιγμή. Σε περίπτωση που η ροή συνοδεύεται και από ηχητική επένδυση η αναπαραγωγή της ηχητικής επένδυσης θα διακόπτεται ταυτόχρονα με τη διακοπή της προβολής.

Πλατφόρμα Υλοποίησης

Η πλατφόρμα που επιλέξαμε για την υλοποίηση της εφαρμογής είναι το λειτουργικό Linux και πιο συγκεκριμένα το Ubuntu (έκδοση για υπολογιστή [51]) της Canonical. Αναφέρουμε σε αυτό το σημείο τους λόγους που μας ώθησαν στην επιλογή του αλλά και τους περιορισμούς που προέκυψαν λόγω αυτού.

Πλεονεκτήματα

Πρόσβαση σε εφαρμογές ανοικτού κώδικα

Μέσω του Linux παρέχεται πρόσβαση σε μια πληθώρα εφαρμογών ανοικτού κώδικα [30] [31] με δημόσια άδεια χρήσης (GNU General Public License [32]). Είδαμε προηγουμένως ότι η επαναχρησιμοποίηση πηγαίου κώδικα και ενθαρρύνεται στα πλαίσια

του Agile άλλα και επιβάλλεται σε περιπτώσεις όπου η απαιτούμενη ανθρωποπροσπάθεια μπορεί να συρρικνωθεί σημαντικά μετατρέποντας ένα έργο το οποίο -χωρίς επαναχρησιμοποίηση κώδικα- δεν θα ήταν εφικτό δεδομένης της ποικιλομορφίας και έκτασής του στα δεδομένα χρονικά πλαίσια. Επίσης πρόκειται για λειτουργικό το οποίο διατίθεται δωρεάν (όπως και οι εφαρμογές με δημόσια άδεια χρήσης) και επομένως διευκολύνει την προοπτική η παρούσα εργασία να συνεχιστεί στα πλαίσια άλλου ερευνητικού ή εκπαιδευτικού έργου ή/και να διαμοιραστεί περαιτέρω για μη εμπορική χρήση χωρίς παραβίαση κάποιας άδειας χρήσης.

Αξιοποίηση Υλικού

Το Linux αξιοποιεί υπολογιστικά συστήματα τόσο μοντέρνας αρχιτεκτονικής και προδιαγραφών όσο και περιορισμένων τεχνικών χαρακτηριστικών. Πρόκειται επομένως για πλατφόρμα που ευνοεί την καλύτερη χρήση των διαθέσιμων υπολογιστικών πόρων και λοιπών χαρακτηριστικών του συστήματος. Επομένως αποφορτίζει μερικώς την υλοποίηση από την ανάγκη να ελέγχεται αυστηρά η κατανάλωση πόρων.

Φορητότητα

Το Linux πλέον χρησιμοποιείται αυτούσιο σε πληθώρα υπολογιστικών συστημάτων (netbook, tablets, laptop, netbook, desktop) και μοιράζεται κοινά στοιχεία (συνήθως σε επίπεδο κελύφους) με άλλα λειτουργικά όπως FreeBSD, το Mac IOS και το Android. Δεν είναι δηλαδή σπάνιο μια εφαρμογή που εκτελείται σε κάποια διανομή Linux να μπορεί να εκτελεστεί με μικρές ή ακόμη και χωρίς μετατροπές σε άλλα λειτουργικά συστήματα. Επομένως ενδείκνυται τόσο από άποψη φορητότητας όσο και ευρείας βάσης χρηστών.

Ευκολία συνένωσης και επικοινωνίας διεργασιών

Κατά τη διάρκεια εξοικείωσης με τα υπάρχοντα συστήματα λογισμικού στα πλαίσια της εργασίας έγινε σαφές ότι θα χρειαζόταν η “συνεργασία” περισσότερων της μιας εφαρμογών για την υλοποίηση του σκέλους “εικόνα” και του σκέλους “ήχος” αντιστοίχως. Το Linux λόγω σχεδίασης παρέχει την ευκολία της συνύπαρξης ετερογενών εφαρμογών και διεργασιών όπως επίσης και την ταυτόχρονη μεταγλώττισή τους και χτισίματος των εκτελέσιμων αρχείων. Αυτό το προσόν διαφάνηκε ως προαπαιτούμενο

για την επαναχρησιμοποίηση των εφαρμογών οι οποίες θα συνιστούσαν τα “δομικά συστατικά” του.

Υιοθέτηση νέων εξελιγμένων τεχνικών ανάπτυξης στις διαθέσιμες εφαρμογές

Η κοινότητα που δραστηριοποιείται στην ανάπτυξη ανοιχτού λογισμικού συχνά προέρχεται από τους κόλπους των ανώτερων και ανώτατων εκπαιδευτικών ιδρυμάτων ή από άλλους τομείς όπου ενθαρρύνονται οι πρωτοποριακές τεχνικές, η χρήση σύγχρονων εργαλείων και γενικότερα ο δημιουργικός πειραματισμός σε πρωτοποριακές λύσεις [33] πέρα από τα τετριμμένα.

Μειονεκτήματα

Έλλειψη αυστηρών προδιαγραφών

Ο ανοιχτός κώδικας, στον οποίο το Linux παρέχει εύκολη πρόσβαση, αποτελεί προϊόν συλλογικής σχεδίασης και συγγραφής. Συχνά, δε, δεν εφαρμόζεται εκτενής έλεγχος στα επιμέρους τμήματα του κώδικα που σταδιακά προστίθενται ή δεν γίνεται εκτενής έλεγχος των συνενωμένων τμημάτων κώδικα ή μπορεί να απουσιάζει εντελώς η τεκμηρίωση του είτε σε μορφή συνοδευτικών εγχειριδίων είτε ακόμη και σε μορφή σχολίων εντός του πηγαίου κώδικα. Αυτό δυσχεραίνει την ανάπτυξη νέων τμημάτων κώδικα και πολλές φορές απαιτεί τη μελέτη του υπάρχοντος κώδικα γραμμή-προς-γραμμή.

Έλλειψη ενιαίας αρχιτεκτονικής και μορφολογίας κώδικα

Η ποικιλομορφία υποδομής των προγραμματιστών που εμπλέκονται στην ανάπτυξη εφαρμογών ανοιχτού κώδικα πολλές φορές οδηγεί και σε ποικιλομορφία στη δομή και αρχιτεκτονική των επιμέρους τμημάτων του. Αυτό λειτουργεί ανασταλτικά στην εύκολη κατανόησή του και κατά συνέπεια στη συντήρηση της υπάρχουσας υλοποίησης και στην εξέλιξή της με ανάπτυξη νέων χαρακτηριστικών.

Υιοθέτηση νέων εξελιγμένων τεχνικών ανάπτυξης στις διαθέσιμες εφαρμογές

Όπως αναφέρθηκε παραπάνω, Η κοινότητα που δραστηριοποιείται στην ανάπτυξη ανοιχτού λογισμικού συχνά προέρχεται από τους κόλπους των ανώτερων και ανώτατων εκπαιδευτικών ιδρυμάτων ή από άλλους τομείς όπου ενθαρρύνονται οι πρωτοποριακές τεχνικές, η χρήση σύγχρονων εργαλείων και γενικότερα ο δημιουργικός πειραματισμός πέρα από τις δοκιμασμένες λύσεις.

Πολυπλοκότητα βημάτων εγκατάστασης εφαρμογών

Συχνά η εγκατάσταση και χρησιμοποίηση κάποιας εφαρμογής απαιτεί περισσότερα βήματα από το “διπλό κλικ” σε κάποιο εκτελέσιμο το οποίο να αναλαμβάνει όλες τις απαραίτητες ενέργειες για τη εγκατάσταση. Κάποιες φορές μπορεί να χρειαστεί πολύωρη ενασχόληση έως ότου επιλυθεί κάποιο πρόβλημα. Αρωγός σε τέτοιου είδους δυσκολίες στέκεται η διαθέσιμη πληροφορία στον παγκόσμιο ιστό ειδικά για τις διανομές Linux με ευρεία βάση χρηστών.

Εργαλεία ανάπτυξης και εκτέλεσης και επαναχρησιμοποίηση υπαρχουσών εφαρμογών

Όπως αναφέρθηκε και νωρίτερα, τόσο τα εργαλεία όσο και οι υπάρχουσες εφαρμογές που χρησιμοποιήθηκαν είναι ανοιχτού κώδικα και διατίθενται δωρεάν για μη-εμπορική χρήση. Η επιλογή εργαλείων ανάπτυξης υπαγορεύθηκε κυρίως από τα κριτήρια επιλογής (μη-λειτουργικές απαιτήσεις) των υπαρχουσών εφαρμογών:

- Ο βαθμός και η έκταση της απαιτούμενης τροποποίησης του υπάρχοντος πηγαίου κώδικα των εφαρμογών δεν ήταν γνωστός εκ των προτέρων. Συνεπώς έπρεπε να γίνει επιλογή εφαρμογών κατά προτίμηση γραμμένων σε κάποια γλώσσα προγραμματισμού με την οποία να υπάρχει ήδη καλό προγραμματιστικό υπόβαθρο. Επίσης υπήρχε η μη-λειτουργική απαίτηση να είναι δυνατή η ενσωμάτωση των δύο εφαρμογών ώστε να μπορούν να μεταγλωττιστούν/χτιστούν αλλά και να εκτελεστούν ως φαινομενικά μία μόνο εφαρμογή.

Για το λειτουργικό σκέλος “διαχείριση δεδομένων ήχου και παραγωγή ηχητικής επένδυσης από αυτά” επιλέχθηκε το Csound το οποίο είναι γραμμένο σε μια ευρέως διαδεδομένη γλώσσα

προγραμματισμού, τη C. Μετά από εκτενή έρευνα δεν βρέθηκε εφαρμογή για τη διαχείριση του σκέλους “διαχείριση φωτογραφίας” γραμμένη στη γλώσσα C και τελικά επιλέχθηκε η εφαρμογή Shotwell [34] η οποία δεν είναι πρωτογενώς γραμμένη σε C όμως ο μεταγλωττιστής της παράγει πηγαίο κώδικα C ως ενδιάμεσο προϊόν (βλ. κεφάλαιο Αποτελέσματα πρακτικής εφαρμογής εργαλείων). Η εφαρμογή Shotwell θα χρειαζόταν να τροποποιηθεί ώστε να υποστηρίζει την εξαγωγή δεδομένων ήχου.

Τα εργαλεία ανάπτυξης που χρησιμοποιήθηκαν ήταν:

- Πλατφόρμα ανάπτυξης λογισμικού Eclipse, έκδοση 3.7.0 “Indigo” [36]. Η συγκεκριμένη πλατφόρμα διαθέτει ολοκληρωμένα περιβάλλοντα ανάπτυξης λογισμικού (IDEs) για διαφορετικές γλώσσες προγραμματισμού. Ωστόσο στα πλαίσια της εργασίας χρησιμοποιήθηκε τελικά μόνον ως επεξεργαστής κειμένου για λόγους οι οποίοι θα αναφερθούν στη συνέχεια.
- Εμπρόσθιο τμήμα (frontend) και επεξεργαστής κειμένου αρχείων εισόδου για το Csound QuteCsound [37]. Παρέχει δυνατότητες επισήμανσης πηγαίου κώδικα Csound, διαδραστικές μικροεφαρμογές (widgets) και δυνατότητα χειρισμού της εγκατεστημένης διανομής του Csound
- Εργαλείο ανάλυσης πρωτοκόλλων δικτύων υπολογιστών Wireshark [38]. Το Wireshark χρησιμοποιήθηκε για την ανάλυση των μηνυμάτων τα οποία ανταλλάσσονται μεταξύ των δύο εφαρμογών (Shotwell και Csound) κατά τη διάρκεια της εκτέλεσης.
- Πρωτόκολλο επικοινωνίας μεταξύ υπολογιστών και ψηφιακών συσκευών παραγωγής ήχου και ηχητικών σημάτων OSC [39]. Το πρωτόκολλο χρησιμοποιήθηκε για την επικοινωνία μεταξύ των εφαρμογών κατά τη διάρκεια εκτέλεσης της εφαρμογής

Οδηγίες για την εγκατάσταση της εφαρμογής παρέχονται στο ΠΑΡΑΡΤΗΜΑ Α.

Αρχιτεκτονική

Το λογισμικό έχει δομηθεί χρησιμοποιώντας τόσο υπάρχουσες λειτουργίες όσο και νέες στις χρησιμοποιούμενες εφαρμογές. Για πληροφορίες σχετικές με την αρχιτεκτονική των επιμέρους εφαρμογών που χρησιμοποιήθηκαν μπορεί κανείς να ανατρέξει στους δικτυακούς τόπους των αντίστοιχων οργανισμών ανάπτυξής τους [44][45]. Οι αλλαγές που πραγματοποιήθηκαν ώστε να καταστεί δυνατή η υλοποίηση του συστήματος της παρούσας εργασίας αφορούν κατά κύριο λόγο στην εφαρμογή Shotwell η οποία δεν διαθέτει κάποια υποδομή εξαγωγής μέσω υπολογισμού δεδομένων από φωτογραφίες. Φυσικά διαθέτει ενσωματωμένη βάση δεδομένων όπου αποθηκεύονται μεταδεδομένα (metadata) από τις φωτογραφίες (λ.χ. ημερομηνία λήψης, μοντέλο φωτογραφικής μηχανής κ.α), όμως αυτά βρίσκονται αποθηκευμένα ήδη ως μεταδεδομένα όταν οι φωτογραφίες εισάγονται στη συλλογή του χρήστη και δεν εξάγονται από την εφαρμογή.

Με τη βοήθεια τροποποιήσεων που έγιναν στο αρχείο το οποίο χρησιμοποιεί ο μεταγλωττιστής Scons (υπεύθυνος για τη μεταγλώττιση του Csound) έγινε δυνατή η ταυτόχρονη μεταγλώττιση και του Shotwell ενώ κατά την εκτέλεση του Csound το πρόγραμμα δημιουργεί μια θυγατρική διεργασία μέσω εντολής διακλάδωσης fork [46]. Στη συνέχεια, το Shotwell εκτελείται από τη νέα θυγατρική διεργασία. Κατά την εκτέλεσή του οι εφαρμογές χρησιμοποιούν. Στη συνέχεια οι δύο εφαρμογές χρησιμοποιούν αμφότερες υποδοχές (sockets) σε επίπεδο πλατφόρμας ώστε να είναι δυνατή η επικοινωνία τους.

Όταν ο χρήστης επιλέξει την προβολή ροής εικόνων με αυτόματη ηχητική επένδυση (βλ. παράρτημα Α για περισσότερες πληροφορίες σχετικά με την εγκατάσταση αλλά και τον τρόπο χρήσης της εφαρμογής), η εφαρμογή Shotwell εξάγει σε πραγματικό χρόνο τα δεδομένα ήχου από τις φωτογραφίες και παράλληλα με την προβολή τους (κατά τη ροή εικόνων) αποστέλλει στο τμήμα της εφαρμογής που είναι υπεύθυνο για την παραγωγή της ηχητικής επένδυσης τα δεδομένα που σχετίζονται με την προβαλλόμενη εικόνα (βλ. παράρτημα Β για πληροφορίες σχετικές με τον τρόπο μετάδοσης των δεδομένων). Ο ρυθμός αποστολής των δεδομένων υπολογίζεται με βάση τη χρονική διάρκεια παραμονής της κάθε φωτογραφίας στην οθόνη του χρήστη κατά τη ροή εικόνων ώστε να διασφαλιστεί ότι η ποσότητα και κατά συνέπεια και η πυκνότητα των δεδομένων

παραμένει σε διαχειρίσιμα επίπεδα.

Η επικοινωνία των επιμέρους εφαρμογών πραγματοποιείται με χρήση του πρωτοκόλλου UDP.

Το OSC δίνει την δυνατότητα επιλογής μεταξύ UDP [48] και TCP [49]. Προτιμήθηκε το UDP για τους παρακάτω λόγους:

- Η απαίτηση υψηλού βαθμού είναι να έχει η εφαρμογή μας χαμηλό χρόνο απόκρισης και το UDP παρέχει τη βέλτιστη λύση όσον αφορά την ταχύτητα σε σχέση με το TCP.

- Η εφαρμογή μας δεν έχει ισχυρή απαίτηση για διαχείριση σφαλμάτων μετάδοσης. Με άλλα λόγια δεν υπάρχει υψηλό ρίσκο σε περίπτωση που η παράδοση κάποιου πακέτου είναι μερικώς ή ολικώς ανεπιτυχής. Συνεπώς δεν χρειάζεται να εξασφαλιστεί η παράδοσή του πακέτου (εγγύηση που παρέχεται από το TCP, εις βάρος όμως της ταχύτητας)

- Τα πακέτα δεδομένων που ανταλλάσσονται κατά την εκτέλεση της εφαρμογής είναι σχετικά πολυάριθμα αλλά μικρά σε μέγεθος που είναι το προφίλ του τύπου επικοινωνίας για την οποία σχεδιάστηκε αρχικά το UDP

- Κατά την κατάργηση της σύνδεσης δεν απαιτούμε να υπάρχει κάποια επικοινωνία μεταξύ των επιμέρους λειτουργικών τμημάτων της εφαρμογής ούτε να υπάρχει σύνδεση μεταξύ τους (κάτι που θα εξασφάλιζε το TCP σε κόστος όμως της ευκολίας υλοποίησης)

Τα πακέτα δεδομένων παραλαμβάνονται από το Csound το οποίο τα αξιοποιεί στα πλαίσια ενός προκαθορισμένου αρχείου εισόδου csd [47] το οποίο περιλαμβάνει κώδικα σύμφωνα με τον οποίο οι ληφθείσες τιμές οδηγούν στην παραγωγή ηχητικής επένδυσης σε πραγματικό χρόνο.

Εξαγωγή δεδομένων από εικόνα στο Shotwell

Όπως προείπαμε, η εφαρμογή Shotwell χρειάστηκε να εμπλουτιστεί και βελτιωθεί ώστε να υποστηρίξει τη δυνατότητα εξαγωγής δεδομένων ήχου. Η μοναδική λειτουργία σχετική με υπολογισμό δεδομένων εικόνας που υφίστατο στην υπάρχουσα εφαρμογή είναι εκείνη που υποστηρίζει την επιλογή του χρήστη να βελτιώσει αυτόματα κάποια

οπτικά στοιχεία της εικόνας όπως την φωτεινότητα, την αντίθεση και τον κορεσμό των χρωμάτων υπό τη μορφή λειτουργίας 'picture enhancement'. Ο υπάρχων κώδικας για να πραγματοποιήσει την βελτιστοποίηση αυτή, εφαρμόζει αλγόριθμους απευθείας στα εικονοστοιχεία της εικόνας και στη συνέχεια αποθηκεύει την εικόνα-προϊόν σε προσωρινό αρχείο στη μνήμη το οποίο μπορεί στη συνέχεια να αποθηκευτεί ή να απορριφθεί.

Για να αποφευχθεί σαν τελικό ηχητικό αποτέλεσμα ο μονοδιάστατος ήχος αξιοποιήσαμε δύο ομάδες δεδομένων εικόνας ως δεδομένα ήχου:

- “Γενικά στοιχεία εικόνας” (φωτεινότητα, κορεσμός και ποσόστωση σε περιοχές χαμηλής, μέσης και υψηλής χρωματικής τονικότητας).
- “Ειδικά/μορφολογικά στοιχεία εικόνας”. Αυτά εξάγονται μέσω του αλγόριθμου ανίχνευσης ακμών του Canny και μαζί με τα γενικά στοιχεία αποτελούν το “ηχητικό αποτύπωμα” της εικόνας.

Όλα τα δεδομένα υπολογίζονται σε πραγματικό χρόνο όταν ο χρήστης επιλέξει την προβολή ροής εικόνων με αυτόματη ηχητική επένδυση. Όπως υπολογίστηκε κατά την περιγραφή λειτουργίας του αλγόριθμου Canny για ανίχνευση ακμών, ο χρόνος εκτέλεσης είναι γραμμική συνάρτηση του αριθμού των εικονοστοιχείων πάνω στα οποία πρέπει να επιδράσει ο αλγόριθμος. Λόγω του γεγονότος ότι πλέον οι αναλύσεις εικόνας που συναντώνται στις φωτογραφίες είναι υψηλές, τέθηκαν τα παρακάτω σχεδιαστικά διλήμματα σχετικά με τη μέθοδο και την χρονική τοποθέτηση της εξαγωγής των δεδομένων από την εικόνα:

1. Η εξαγωγή του ηχητικού αποτυπώματος πραγματοποιείται κατά την εισαγωγή της φωτογραφίας στις βιβλιοθήκες του χρήστη. Στη συνέχεια το ηχητικό αποτύπωμα αποθηκεύεται ώστε να είναι διαθέσιμο σε περίπτωση που ο χρήστης επιλέξει την προβολή ροής εικόνας με ηχητική επένδυση (εξαγωγή δεδομένων σε “νεκρό” χρόνο-offline προσέγγιση).
2. Η εξαγωγή του ηχητικού αποτυπώματος πραγματοποιείται μόνο σε περίπτωση που ο χρήστης επιλέξει την προβολή ροής εικόνας με ηχητική επένδυση σε πραγματικό χρόνο (εξαγωγή δεδομένων όταν αυτά ζητηθούν – online/”lazy start” προσέγγιση)

Παρακάτω παρατίθενται τα πλεονεκτήματα και τα μειονεκτήματα των δύο αυτών προσεγγίσεων:

Εξαγωγή ηχητικού αποτυπώματος με την offline προσέγγιση

Πλεονεκτήματα

-Μπορεί να γίνει πλήρης απόκρυψη του επιπλέον χρόνου εκτέλεσης από τον χρήστη ασχέτως από τον βαθμό πολυπλοκότητας των υπολογισμών και το μέγεθος της φωτογραφίας

-Η κατανάλωση υπολογιστικών πόρων για την εξαγωγή του ηχητικού αποτυπώματος πραγματοποιείται σε φάση όπου δεν καταναλώνονται πόροι για την αναπαραγωγή της ηχητικής επένδυσης

Μειονεκτήματα

-Επιφόρτιση του συστήματος με περιττούς υπολογισμούς σε περίπτωση που ο χρήστης δεν επιθυμεί τελικά να χρησιμοποιήσει την επιλογή “ροή με ηχητική επένδυση”

-Δημιουργία ανάγκης αποθήκευσης των εξαχθέντων δεδομένων στον σκληρό δίσκο. Αυτό συνεπάγεται την ανάγκη για πληθώρα επιπλέον σχεδιαστικών αποφάσεων, όπως για παράδειγμα το πως θα πραγματοποιείται η αποθήκευση των δεδομένων. Μερικές εναλλακτικές λύσεις σε αυτό το πρόβλημα για παράδειγμα θα ήταν:

(α) Ως μεταδεδομένο μαζί με το αρχείο της φωτογραφίας.

(β) Ως αυτόνομη καταχώρηση στη βάση δεδομένων.

-Δημιουργία ανάγκης επικαιροποίησης (update) και ανάκτησης των δεδομένων ήχου όταν κάποιο στοιχείο της φωτογραφίας μεταβληθεί (π.χ φωτεινότητα, κορεσμός χρωμάτων).

-Επιπλέον έγινε αντιληπτό κατόπιν μελέτης του υπάρχοντος σχεδιασμού του συστήματος ότι η offline προσέγγιση θα απαιτούσε εκτενή επανασχεδιασμό της αρχιτεκτονικής του συστήματος. Ο λόγος ήταν ότι η πρόσβαση στα εικονοστοιχεία της εικόνας είναι δυνατή μόνο μετά το τέλος της διαδικασίας εισαγωγής της.

Εξαγωγή δεδομένων με την online προσέγγιση

Πλεονεκτήματα

-Οι υπολογισμοί πραγματοποιούνται μόνο κατόπιν αιτήσεως και επομένως δεν έχουμε σπατάλη υπολογιστικών πόρων

-Η αξιοποίηση των πόρων του συστήματος γενικότερα βελτιστοποιείται μιας και η εξαγωγή του ηχητικού αποτυπώματος γίνεται χρησιμοποιώντας το στιγμιότυπο της εικόνας που είναι διαθέσιμο τη χρονική στιγμή της προβολής τη ροής εικόνων. Επομένως τυχόν αλλαγές που έχουν προηγηθεί στα στοιχεία της εικόνας έχουν ήδη ενσωματωθεί σε αυτή και δεν απαιτείται επανάληψη υπολογισμού.

-Καταργείται η ανάγκη για αποθήκευση των δεδομένων στο σκληρό δίσκο μαζί με όλα τα συνεπαγόμενα της.

-Ελαχιστοποιείται η ανάγκη επέμβασης στη υπάρχουσα λογική σχεδίαση της εφαρμογής

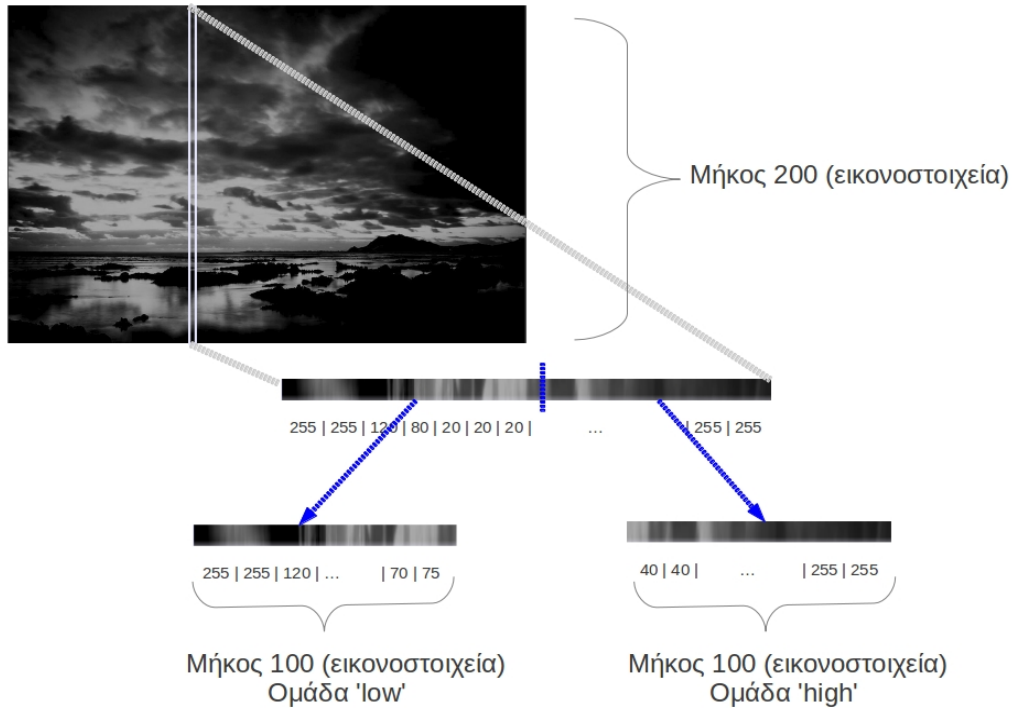
Μειονεκτήματα

-Μπορεί να οδηγήσει σε ορατή από το χρήστη υστέρηση κατά την προβολή της ροής

-Μπορεί να οδηγήσει σε υψηλή -τοπικά- ανάγκη για υπολογιστικούς πόρους μιας και η εξαγωγή των δεδομένων θα πραγματοποιείται παράλληλα με την παραγωγή της ηχητικής επένδυσης η οποία επίσης έχει ανάγκες σε υπολογιστικούς πόρους.

Επιλέχθηκε η online προσέγγιση για την οποία όμως χρειάστηκε να βρεθεί κάποιο σχεδιαστικό “τέχνασμα” ώστε να ελαχιστοποιηθεί η επίδραση στον χρόνο απόκρισης του συστήματος. Το τέχνασμα αυτό έχει να κάνει με την παραδοχή ότι τα γεωμετρικά στοιχεία της φωτογραφίας δεν επηρεάζονται δραστικά από το μέγεθος της φωτογραφίας. Είναι δε σημαντικό να αναφέρουμε ότι ο ανθρώπινος εγκέφαλος αντιλαμβάνεται το σημασιολογικό περιεχόμενο ακόμη και φωτογραφιών σε ανάλυση 32x32 [50]. Η τακτική που ακολουθήσαμε, πριν από την εξαγωγή των δεδομένων ήχου, ελαττώνει το μέγεθος της εικόνας ώστε αυτή να μην υπερβαίνει τα 40000 εικονοστοιχεία. Η επιλογή αυτού του μεγέθους οδήγησε σε χαμηλούς χρόνους (<250 milliseconds) εξαγωγής δεδομένων ήχου ανά φωτογραφία. Και συνεπάγεται ότι το προφίλ το οποίο συνολικά αποστέλλεται μέσω των μηνυμάτων OSC (στην περίπτωση αποστολής ειδικών στοιχείων) έχει μήκος συνολικά ίσο με 200 που αντιστοιχεί στο ύψος (εκφρασμένου σε αριθμό εικονοστοιχείων) της εικόνας μετά την ελάττωση του μεγέθους. Δεδομένα για το πλήρες μήκος αποστέλλονται με χρήση 2 μηνυμάτων λόγω περιορισμού που επιβλήθηκε

από το μέγιστο μέγεθος μηνύματος που μπορεί να χειριστεί η υλοποίηση του OSC στο Csound.



Εικόνα 19 Ο κατακερματισμός του προφίλ εικόνας σε δύο τμήματα για αποστολή σε αντίστοιχα OSC μηνύματα

Το Shotwell αποστέλλει το ηχητικό αποτύπωμα μέσω μηνυμάτων OSC στη UDP θύρα 7613. Χρησιμοποιούνται δύο διαφορετικά μηνύματα για την αποστολή των γενικών και των ειδικών στοιχείων της φωτογραφίας αντιστοίχως και η διάκριση των μηνυμάτων γίνεται βάσει ένδειξης που επίσης αποστέλλεται με το μήνυμα. Στο παράρτημα Β φαίνονται μερικά μηνύματα OSC στα οποία περιέχονται δεδομένα ήχου και που αποστέλλονται από το Shotwell σε πραγματικό χρόνο. Αξίζει να σημειωθεί ότι μεταξύ των άλλων δεδομένων που μεταφέρονται στη συνεργαζόμενη εφαρμογή περιλαμβάνεται

και μια τιμή κινητής υποδιαστολής (float) η οποία αντιπροσωπεύει τη διάρκεια του ηχητικού σήματος. Σημαντικό επίσης χαρακτηριστικό της επικοινωνίας μεταξύ του Shotwell και της συνεργαζόμενης εφαρμογής είναι το ότι τα ειδικά δεδομένα (τα οποία αντιστοιχούν σε κάθετες τομές/προφίλ ύψους).

Χρησιμοποιείται επίσης η UDP θύρα 7614 για την αποστολή μηνυμάτων ελέγχου προς τη συνεργαζόμενη εφαρμογή. Μέσω των μηνυμάτων ελέγχου μπορεί να ζητηθεί από τη συνεργαζόμενη εφαρμογή να διακόψει ή να επανεκκινήσει την αναπαραγωγή της ηχητικής επένδυσης.

Διαχείριση δεδομένων ήχου στο Csound

Το Csound διαθέτει ένας πλήθος δυνατοτήτων για την παραγωγή ήχου. Είναι, δε, δημοφιλές στους χώρους παραγωγής ηλεκτρονικής μουσικής για τη δυνατότητα που προσφέρει να λειτουργεί τόσο ως αυτόνομη οντότητα όσο και ως συμπληρωματική εφαρμογή σε άλλα συστήματα λογισμικού ή συσκευές παραγωγής ψηφιακών ηχητικών σημάτων όπως για παράδειγμα άλλες εφαρμογές που εκτελούνται στον ίδιο ή σε άλλο υπολογιστή ο οποίος επικοινωνεί μέσω δικτύου με τον υπολογιστή που φιλοξενεί το Csound, midi ή ηλεκτρονικά μουσικά όργανα. Κατά την αυτόνομη λειτουργία του Csound, ο χρήστης μπορεί να κατασκευάσει μεμονωμένους ήχους ή και ολοκληρωμένα ηχητικά/μουσικά θέματα χρησιμοποιώντας τη γλώσσα προγραμματισμού του Csound η οποία προορίζεται για την παραγωγή ήχου. Ενώ για την συνεργασία του Csound με άλλες συσκευές απαιτείται η χρήση του πρωτοκόλλου OSC ώστε το Csound να είναι σε θέση να επικοινωνεί με τις συσκευές αυτές. Ο προγραμματισμός του Csound, μεταξύ άλλων μπορεί να επιτευχθεί με χρήση αρχείων εισόδου τύπου csd (παράδειγμα της δομής τέτοιου αρχείου μπορεί να βρεθεί στον σύνδεσμο <http://csounds.com/manual/html/examples/oscils.csd>).

Η βασική δομή του csd αρχείου περιλαμβάνει τα παρακάτω πεδία:

- CsOptions όπου περιέχονται ρυθμίσεις καθολικής εμβέλειας
- CsInstruments όπου ορίζονται τα “όργανα” τα οποία αξιοποιούνται για την παραγωγή του ηχητικού θέματος
- CsScore όπου περιλαμβάνεται ο πίνακας βάσει του οποίου καθορίζεται η εκτέλεση του μουσικού θέματος.

Η γλώσσα προγραμματισμού Csound διαθέτει έναν μακρύ κατάλογο έτοιμων μακροεντολών/συναρτήσεων οι οποίες ονομάζονται κωδικοί ενέργειας (opcodes) (<http://www.csounds.com/manual/html/MiscQuickref.html>) αλλά και πολυάριθμους τελεστές. Πολλοί από τους διαθέσιμους τελεστές είναι κοινοί με άλλες γλώσσες προγραμματισμού, όπως λ.χ. τελεστές δυαδικής λογικής ενώ άλλοι είναι αποκλειστικοί της εφαρμογής.

Για τις ανάγκες της εφαρμογής στα πλαίσια της εργασίας εκμεταλλευθήκαμε τη δυνατότητα λήψης, από το Csound, δεδομένων από εξωτερικές πηγές και χρησιμοποιήσαμε τα παρακάτω στοιχεία αρχιτεκτονικής τα οποία υλοποιήθηκαν μέσω csd αρχείου:

- Είναι δυνατή η λήψη μηνυμάτων σε δύο UDP θύρες: την 7613 και 7614.
- Η θύρα 7613 χρησιμοποιείται για τη λήψη μηνυμάτων που μεταφέρουν το ηχητικό αποτύπωμα. Διακρίνονται, με χρήση εσωτερικής ένδειξης, σε εκείνα που μεταφέρουν τα γενικά και σε εκείνα που μεταφέρουν τα ειδικά στοιχεία αντιστοίχως. Ενώ εκείνα που μεταφέρουν τα ειδικά δεδομένα διακρίνονται περαιτέρω σε χαμηλό (low) και υψηλό (high) τμήμα του προφίλ.
- Η θύρα 7614 χρησιμοποιείται για τη λήψη μηνυμάτων ελέγχου τα οποία μεταφέρουν εντολή διακοπής ή επανεκκίνησης της ηχητικής αναπαραγωγής
- Η σύνθεση των ήχου χρησιμοποιεί ως είσοδο τα δεδομένα που λαμβάνονται μέσω των μηνυμάτων ενώ η υλοποίηση επιτρέπει τη διαφοροποίηση στο χειρισμό δεδομένων με κριτήριο το αν αυτά αφορούν τα γενικά ή ειδικά στοιχεία.
- Όσον αφορά τα ειδικά δεδομένα ο τρόπος διαχείρισης τους για την παραγωγή ήχου λαμβάνει υπόψη όχι μόνο την περιεχόμενη τιμή αλλά επίσης το ύψος στο

οποίο αντιστοιχεί την κατανομή σύμφωνα με τη λογική που απεικονίζεται στην του προηγούμενου κεφαλαίου και το αν πρόκειται για την ομάδα 'high' ή την ομάδα 'low'.

- Τα δεδομένα που λαμβάνονται μέσω του OSC μηνύματος δεν αντιστοιχίζονται ένα προς ένα σε διαφορετικής συχνότητας ήχους προς αποφυγή δημιουργίας θορύβου. Αντιθέτως οι τιμές που αντιστοιχούν σε ένα γειτονικό εύρος ομαδοποιούνται και αξιοποιούνται ομαδικά για την παραγωγή πιο αρμονικού αποτελέσματος.

Στη συνέχεια δίδονται μερικά από τα opcodes και τους τελεστές που χρησιμοποιήθηκαν για την παραγωγή της ηχητικής επένδυσης καθώς και η σημασία τους για τη συγκεκριμένη υλοποίηση όπου αυτό παρουσιάζει ενδιαφέρον.

sr: θέτει το ρυθμό δειγματοληψίας ανά δευτερόλεπτο ανά κανάλι ήχου. Χρησιμοποιήθηκε τιμή χαμηλότερη της προεπιλεγμένης ώστε να μπορούμε να χειριστούμε μεγαλύτερο αριθμό σημάτων δεδομένου επίσης ότι ο υπολογιστής ο οποίος χρησιμοποιήθηκε για την ανάπτυξη και τον έλεγχο του συστήματος λογισμικού δεν διέθετε προηγμένη κάρτα ήχου.

ksmps: θέτει τον αριθμό δειγματοληψιών ανά περίοδο ελέγχου

nchnls: καθορίζει τον αριθμό των καναλιών που χρησιμοποιούνται για την παραγωγή του ήχου

OSCinit: πρόκειται για την εντολή που θέτει την εφαρμογή σε ετοιμότητα λήψης μηνυμάτων σε κάποια θύρα που καθορίζεται μέσω της εντολής. Στην περίπτωσή μας χρησιμοποιήσαμε τις θύρες 7613 και 7614.

OSClisten: στο σημείο αυτό γίνεται η λήψη μηνυμάτων OSC. Χρησιμοποιήθηκαν διαφορετικές τέτοιες μακροεντολές για τα διαφορετικού τύπου OSC Μηνύματα που

υποστηρίζονται στα πλαίσια της εφαρμογής.

alwayson: καθορίζει το αν κάποιο όργανο θα παραμένει ενεργοποιημένο χωρίς να χρειάζεται να ενεργοποιηθεί δυναμικά κατά τη διάρκεια εκτέλεσης.

event: δημιουργεί ένα συμβάν μέσα από όργανο. Σε αντίθεση με την προκαθορισμένη αλληλουχία συμβάντων η οποία καθορίζεται από το πεδίο CsScore του αρχείου csd, ο κωδικός ενέργειας event κάνει δυνατή την ενεργοποίηση ενός οργάνου μέσω ενός άλλου. Για παράδειγμα, το όργανο instr1 μπορεί μέσω της ενέργεια event να ενεργοποιήσει την κλήση του οργάνου instrClarinet. Το opcode αυτό χρησιμοποιήθηκε εντατικά ώστε να είναι δυνατή η εκτέλεση διαφορετικών οργάνων μέσω ενός κεντρικού οργάνου το οποίο περιείχε το opcode OSClisten μιας και το τελευταίο δε μπορεί να τοποθετηθεί κεντρικά. Δηλαδή δε μπορεί να γίνεται λήψη μηνυμάτων κεντρικά αλλά μόνο εντός ορισμένου οργάνου.

limit: καθορίζει το άνω και κάτω όριο για ένα εύρος τιμών μεταβλητής.

opcode: δίνει τη δυνατότητα στο χρήστη να καθορίσει ένα νέο opcode.

tone: χρησιμοποιείται για τη δημιουργία ενός χαμηλοδιαβατού φίλτρου, δηλαδή ενός φίλτρου το οποίο αποκόπτει συχνότητες μεγαλύτερες από κάποιο καθορισμένο κατώφλιο. Χρησιμοποιήθηκε για περιορισμό συχνοτήτων σε περιοχές όπου επιτυγχάνεται μεγαλύτερη αρμονία.

turnoff: χρησιμοποιείται για την αποσιώπηση άλλων οργάνων κατά την εκτέλεση. Χρησιμοποιήθηκε για να υλοποιηθεί η σίγαση της ηχητικής επένδυσης όταν ο χρήστης επιλέγει να διακόψει την προβολή της ροής εικόνων με ηχητική επένδυση.

6 ΑΠΟΤΕΛΕΣΜΑΤΑ

Διαδικασία Δοκιμών

Οι δοκιμές του λογισμικού έγιναν στον υπολογιστή όπου πραγματοποιήθηκαν και όλα τα προηγούμενα στάδια σχεδιασμού και ανάπτυξης, έναν φορητό υπολογιστή (laptop) Sony Vaio με τα παρακάτω χαρακτηριστικά:

Επεξεργαστής: Intel Core i5 CPU στα 2.53GHz

Μνήμη: 3909MB

Λειτουργικό σύστημα : Ubuntu 11.10

Κάρτα Γραφικών: ATI Technologies Inc Madison [AMD Radeon HD 5000M Series]

Κάρτα Ήχου:ATI Technologies Inc Redwood HDMI Audio [Radeon HD 5600 Series]

Δοκιμές πραγματοποιήθηκαν τόσο στην πλευρά του Shotwell όσο και στην πλευρά του Csound προτού να γίνει συνολική επαλήθευση της εφαρμογής. Για το Shotwell συναντήθηκαν κατά την αποσφαλμάτωση αρκετές δυσκολίες οι οποίες αναφέρονται αναλυτικά στο κεφάλαιο **Εργαλεία ανάπτυξης και εκτέλεσης και επαναχρησιμοποίηση υπαρχουσών εφαρμογών**. Για το λειτουργικό τμήμα της εφαρμογής που υλοποιείται από το Csound οι δοκιμές πραγματοποιήθηκαν κυρίως μέσω της διαδικασίας δοκιμής-λάθους (trial-and-error) χρησιμοποιώντας είσοδο η οποία δημιουργήθηκε είτε με προσομοίωση (στέλνοντας δεδομένα τα οποία είχα κατασκευαστεί χειροκίνητα) είτε από το Shotwell χρησιμοποιώντας πραγματικές αλλά και τεχνητά κατασκευασμένες φωτογραφίες όπως για παράδειγμα οι παρακάτω:



Εικόνα 20 Δείγματα εικόνας ελέγχου

Επίσης δοκιμάστηκαν φωτογραφικά αρχεία διαφορετικών μεγεθών για να επαληθευτούν οι χρόνοι εκτέλεσης αλλά και να ελεγχθεί ο κώδικας μετατροπής μεγέθους των φωτογραφιών.

Αποτελέσματα πρακτικής εφαρμογής εργαλείων

Κρίνεται σκόπιμο σε αυτή τη θέση να περιγράψουμε κάποιες από τις ιδιαιτερότητες του περιβάλλοντος ανάπτυξης και της αποσφαλμάτωσης του κώδικα που προέκυψαν λόγω της ιδιομορφίας της ανάπτυξης του λογισμικού με επαναχρησιμοποίηση κώδικα. Η περιγραφή αυτή παρουσιάζει ενδιαφέρον κυρίως ως καταγραφή πιθανών προβλημάτων που μπορούν να προκύψουν κατά την υιοθέτηση πρωτοποριακών τεχνικών και εργαλείων τα οποία βρίσκονται σε αρχικό ή πειραματικό ακόμη στάδιο. Κρίνεται, δε, επωφελής μιας και αυτή η στρατηγική αποτελεί συχνά μέρος της Agile φιλοσοφίας ανάπτυξης λογισμικού. Όπως έχει ήδη αναφερθεί, στα πλαίσια της επαναχρησιμοποίησης πηγαίου κώδικα επιλέξαμε τις εφαρμογές Shotwell και Csound. Για το τελευταίο, ο όρος “εφαρμογή” είναι καταχρηστικός. Στην πραγματικότητα είναι μεταγλωττιστής της γλώσσας προγραμματισμού Csound η οποία είναι βελτιστοποιημένη για παραγωγή ήχου και μουσικής, αλλά θα αναφερόμαστε με τον όρο “εφαρμογή”. Τα εκτελέσιμα και των δύο προκύπτουν από τη μεταγλώττιση της γλώσσας προγραμματισμού C. Ωστόσο στην περίπτωση του Shotwell, η εφαρμογή δεν είναι πρωτογενώς γραμμένη στη γλώσσα C

αλλά σε μια μοντέρνα αντικειμενοστραφή γλώσσα προγραμματισμού τη Vala [40] η οποία έχει συντακτικό παρεμφερές με εκείνο της C# και γνωρίζει τα τελευταία χρόνια ιδιαίτερη άνθηση. Ο μεταγλωττιστής Valac της Vala αναλαμβάνει την παραγωγή κώδικα C για τον οποίο στη συνέχεια μπορεί να χρησιμοποιηθεί οποιοσδήποτε C μεταγλωττιστής. Έτσι ο αναλυτής/προγραμματιστής, ενώ συγγράφει πηγαίο κώδικα σε μια σύγχρονη και εξελιγμένη αντικειμενοστραφή γλώσσα προγραμματισμού, επωφελείται από τον πλούτο των βιβλιοθηκών της C αλλά και από ταχύτητα εκτέλεσης του κώδικα εφάμιλλη με εκείνη των εφαρμογών οι οποίες είναι πρωτογενώς γραμμένες σε C. Η σύλληψη ακούγεται (και είναι) σημαντική όσο και πρωτοποριακή. Ωστόσο η Vala πάσχει αρκετά όσον αφορά τον τομέα της αποσφαλμάτωσης (debugging) του κώδικα. Οποιοσδήποτε έχει εργαστεί με έργα γραμμένα στη γλώσσα C γνωρίζει τις δυνατότητες που παρέχονται από ένα εργαλείο αποσφαλμάτωσης όπως είναι το λογισμικό αποσφαλμάτωσης ανοικτού κώδικα του GNU Project (GNU Project Debugger, GDB) [41].

Περίληπτικά η ισχύς του εργαλείου συνίσταται στη δυνατότητα που παρέχει στον προγραμματιστή να επιβλέπει, αναλύει και να τροποποιεί την εκτέλεση του λογισμικού. Στα πλαίσια μιας απλής διαδικασίας αποσφαλμάτωσης, ο GDB, δίνει τη δυνατότητα στον προγραμματιστή να σταματήσει την εκτέλεση του προγράμματος σε επιλεγμένα σημεία εκτέλεσης (breakpoints) και να εξετάσει την ορθότητα των τιμών όλων των χρησιμοποιούμενων μεταβλητών/καταχωρητών αλλά και την ορθότητα της ροής εκτέλεσης του προγράμματος.

```
File Edit View Terminal Help
stm32f10x_rcc.c
312
313 /* Check the parameters */
314 assert(IS_RCC_SYSCLK_SOURCE(RCC_SYSCLKSource));
315
316 tmpreg = RCC->CFGR;
317
318 /* Clear SW[1:0] bits */
B+> 319 tmpreg &= CFGR_SW_Mask;
320
321 /* Set SW[1:0] bits according to RCC_SYSCLKSource value */
322 tmpreg |= RCC_SYSCLKSource;
323
324 /* Store the new value */
325 RCC->CFGR = tmpreg;
326 }
327

0x322 <RCC_SYSCLKConfig+2> sub sp, #20
0x324 <RCC_SYSCLKConfig+4> add r7, sp, #0
0x326 <RCC_SYSCLKConfig+6> str r0, [r7, #4]
0x328 <RCC_SYSCLKConfig+8> mov.w r3, #0
0x32c <RCC_SYSCLKConfig+12> str r3, [r7, #12]
0x32e <RCC_SYSCLKConfig+14> ldr r3, [pc, #40] ; (0x358 <RCC_SYSCLKConfig+56>)
0x330 <RCC_SYSCLKConfig+16> ldr r3, [r3, #4]
0x332 <RCC_SYSCLKConfig+18> str r3, [r7, #12]
B+> 0x334 <RCC_SYSCLKConfig+20> ldr r3, [r7, #12]
0x336 <RCC_SYSCLKConfig+22> bic.w r3, r3, #3
0x33a <RCC_SYSCLKConfig+26> str r3, [r7, #12]
0x33c <RCC_SYSCLKConfig+28> ldr r2, [r7, #12]
0x33e <RCC_SYSCLKConfig+30> ldr r3, [r7, #4]
0x340 <RCC_SYSCLKConfig+32> orr.w r3, r2, r3
0x344 <RCC_SYSCLKConfig+36> str r3, [r7, #12]
0x346 <RCC_SYSCLKConfig+38> ldr r2, [pc, #16] ; (0x358 <RCC_SYSCLKConfig+56>)

RCC_SYSCLKConfig
Arglist at 0x200007f0, args:
Locals at 0x200007f0, Previous frame's sp is 0x200007f8
Saved registers:
r7 at 0x200007f0, lr at 0x200007f4
(gdb) stepi
(gdb) si
RCC_SYSCLKConfig (RCC_SYSCLKSource=0) at stm32f10x_rcc.c:310
(gdb) si
(gdb) si
(gdb) si
(gdb) si
(gdb) b *0x334
Breakpoint 1 at 0x334: file stm32f10x_rcc.c, line 319.
(gdb) c
Continuing.

Breakpoint 1, RCC_SYSCLKConfig (RCC_SYSCLKSource=0) at stm32f10x_rcc.c:319
(gdb)
```

Εικόνα 211 Αποσφαλμάτωση με το GDB

Κρίσιμης σημασίας είναι και η χρήση του συγκεκριμένου λογισμικού αποσφαλμάτωσης μετά από κατάρρευση (crash) του εκτελέσιμου.

Κατά την αποσφαλμάτωση όμως ενός προγράμματος γραμμένου σε Vala δεν είναι πάντοτε δυνατή η πρόσβαση σε όλες τις μεταβλητές του Vala κώδικα. Αυτό συμβαίνει διότι ο κώδικας C που τελικά παράγεται, χρησιμοποιεί μια πληθώρα διαφορετικών τοπικών μεταβλητών με πολυάριθμες εσωτερικές αναθέσεις τιμών οι οποίες δεν είναι εύκολα προσβάσιμες. Επίσης χρησιμοποιεί διαφορετική ονοματολογία συναρτήσεων σε σχέση με τις “Vala-ομόλογές” τους. Παρακάτω φαίνεται ένα απόσπασμα C κώδικα όπως παράγεται από τον μεταγλωττιστή valac

```
5882 #line 684 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5883     tmp8 = self->priv->developments;
5884 #line 684 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5885     g_static_rec_mutex_unlock (&self->priv->__lock_developments);
5886 #line 5887 "Photo.c"
5887 }
5888 #line 689 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5889     return;
5890 #line 5891 "Photo.c"
5891 }
5892 #line 692 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5893     photo_revert_to_master (self, FALSE);
5894 #line 695 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5895     tmp9 = self->row;
5896 #line 695 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5897     tmp10 = d;
5898 #line 695 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5899     tmp9->developer = tmp10;
5900 #line 696 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5901     tmp11 = self->priv->developments;
5902 #line 696 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5903     tmp12 = d;
5904 #line 696 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5905     tmp13 = gee_abstract_map_get (GEE_ABSTRACT_MAP (tmp11), GINT_TO_POINTER (tmp12));
5906 #line 696 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5907     backing_photo_row_unref0 (self->backing_photo_row);
5908 #line 696 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5909     self->backing_photo_row = (BackingPhotoRow*) tmp13;
5910 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5911     tmp14 = self->priv->readers;
5912 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5913     tmp15 = self->backing_photo_row;
5914 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5915     tmp16 = tmp15->file_format;
5916 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5917     tmp17 = self->backing_photo_row;
5918 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5919     tmp18 = tmp17->filepath;
5920 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5921     tmp19 = photo_file_format_create_reader (tmp16, tmp18);
5922 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5923     photo_file_adapter_unref0 (tmp14->developer);
5924 #line 697 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5925     tmp14->developer = tmp19;
5926 #line 699 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5927     tmp20 = self->backing_photo_row;
5928 #line 699 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5929     tmp21 = tmp20->original_orientation;
5930 #line 699 "/home/mike/EAP/Diplomatikh/soundtrackgen/Shotwell/shotwell-0.12.3/src/Photo.vala"
5931     photo_file_adapter_unref0 (tmp14->developer);
```

Εικόνα 222 Δείγμα C κώδικα που παράγει ο μεταγλωττιστής valac

Έτσι δυσχεραίνεται σε σημαντικό βαθμό το απαραίτητο στάδιο της αποσφαλμάτωσης και τελικά αυτή υποβαθμίζεται σε χρονοβόρα και επίπονη διαδικασία συνδυασμού προσθήκης πολυάριθμων εντολών εκτύπωσης εντός του κώδικα (και διαδοχικών μεταγλωττίσεων) και υποθέσεων/βολιδοσκοπήσεων για πιθανές πηγές λαθών.

Διαφορετικής μορφής προκλήσεις αντιμετωπίστηκαν λόγω της ανάγκης χρήσης του πρωτοκόλλου OSC στην υλοποίηση μας μιας και ήταν απαραίτητο για την επικοινωνία μεταξύ των δύο περιεχομένων εφαρμογών. Η πρώτη πρόκληση προέκυψε από το γεγονός ότι ώστε να υιοθετηθεί η χρήση του OSC από την εφαρμογή Shotwell απαιτήθηκε να γίνει δυναμική σύνδεση του Shotwell με τη βιβλιοθήκη liblo [42] (υλοποίηση του OSC σε μορφή βιβλιοθήκης). Για την ενσωμάτωση, όμως, της βιβλιοθήκης liblo χρειάστηκε πρώτα να δημιουργηθεί μια διεπαφή προγραμματισμού εφαρμογών (Application Programming Interface - API) η οποία να κάνει δυνατή τη χρήση από το Shotwell (γραμμένο σε Vala) της βιβλιοθήκης (της οποίας η υλοποίηση είναι σε C) [43]. Στην πλευρά του Csound το OSC υποστηρίζεται ήδη. Ωστόσο παρατηρήθηκε κατά τη διάρκεια της ανάπτυξης ότι η υλοποίηση του OSC στο Csound δεν υποστηρίζει έναν συγκεκριμένο τύπο δεδομένων (με το γλαφυρό όνομα blob) εντός του OSC μηνύματος και ο οποίος παρέχει τη δυνατότητα μετάδοσης μεγάλου όγκου δεδομένων μέσω ενός μόνο OSC μηνύματος. Αυτός ο περιορισμός οδήγησε σε μια συμβιβαστική λύση κατά την οποία χρησιμοποιήθηκαν, εναλλακτικώς, δεδομένα μήκους 32 δυφίων με τη συμβολή απλοϊκής κωδικοποίησης/συμπίεσης ώστε να περιληφθεί όλη η απαιτούμενη πληροφορία προς μετάδοση από το Shotwell προς το Csound. Επίσης, για τον ίδιο λόγο, η αποστολή των δεδομένων εικόνας πραγματοποιείται σε δυο στάδια για κάθε προφίλ εικόνας που αποστέλλεται προς το Csound.

Ευρήματα κατά την εκτέλεση

Το βασικό πρόβλημα που αντιμετωπίστηκε κατά την ανάπτυξη και εκτέλεση πηγάζει στο ότι η επικοινωνία μέσω του πρωτοκόλλου OSC στο Csound δεν προορίζεται για τον

τρόπο με τον οποίο χρησιμοποιείται στο σύστημα STGen το οποίο αναπτύχθηκε στα πλαίσια της εργασίας. Η συνήθης εφαρμογή του OSC στο Csound έγκειται στον έλεγχο ενός οργάνου ή περισσότερων Csound οργάνων μέσω κάποιας εξωτερικής συσκευής. Δηλαδή το Csound υπο φυσιολογικές συνθήκες καλείται να επεξεργαστεί στα πλαίσια της εκτέλεσης ενός οργάνου ένα γραμμικό σήμα που λαμβάνεται μέσω της λειτουργίας OSClisten. Δεν είναι άλλωστε τυχαίο ότι η OSClisten δε μπορεί να τοποθετηθεί εκτός του ορισμού κάποιου οργάνου (instr) στα αρχεία προγραμματισμού του Csound. Αυτός ο περιορισμός οδηγεί κάποιες φορές και σε αποκοπή του ήχου ή αντιαισθητικούς ήχους “κλικ” κατά την αναπαραγωγή της ηχητικής επένδυσης.

Ένα άλλο συμπέρασμα που προέκυψε από την εκτέλεση είναι ότι η πυκνότητα οπτικής πληροφορίας που περιέχεται σε μια συνηθισμένη (μη κατασκευασμένη/επεξεργασμένη) φωτογραφία δεν είναι εφικτό να αποδοθεί ακουστικά στο σύνολό της. Χρειάζεται να γίνουν μερικές παραδοχές και συμβιβασμοί όσον αφορά το ποιο υποσύνολο -τελικά- δεδομένων της εικόνας θα αποδοθούν ακουστικά. Στην περίπτωση του STGen μέρος μόνο όλων των προφίλ της φωτογραφίας αποστέλλεται για παραγωγή ήχου μιας και τα ηχητικά συμβάντα που θα μπορούσαν να προκληθούν στην αντίθετη περίπτωση θα ήταν εξαιρετικά μεγάλης πυκνότητας για να μπορούν να είναι διαχειρίσιμα.

Αποτελέσματα Χρήσης

Κατά την παρατήρηση της επαφής των χρηστών με την εφαρμογή παρατηρήθηκαν τα παρακάτω:

- Ο βαθμός χρηστικότητας του STGen φαίνεται να είναι αρκετά υψηλός απαιτώντας ελάχιστη προσπάθεια από το μέσο χρήστη. Αυτό φυσικά οφείλεται σε καταλυτικό βαθμό στην προϋπάρχουσα υλοποίηση του Shotwell αλλά ενδεχομένως συντέλεσε θετικά και η προσέγγιση που ακολουθήθηκε κατά το σχεδιασμό και την υλοποίηση για ελάχιστη ανάγκη παρέμβασης (και σκέψης) από την πλευρά του χρήστη κατά τη λειτουργία [52].

- Η περιγραφή των απαιτούμενων βημάτων για την “εγκατάσταση” της εφαρμογής μέσω γραμμής εντολών φάνηκε να «τρομάζει» τους μέσους –μη εξοικειωμένους με λειτουργικά συστήματα Linux- χρήστες. Αυτό όμως οφείλεται εν μέρει και στο γεγονός ότι στο Linux γενικώς, η εγκατάσταση κάποιας εφαρμογής, όταν δεν πραγματοποιείται αυτοματοποιημένα από το σύστημα, απαιτεί συνήθως κάπως μεγαλύτερη ενεργό συμμετοχή από ότι στα πιο δημοφιλή λειτουργικά. Ωστόσο στάθηκε δυνατή η απλούστευση της χρήσης της εφαρμογής (κατόπιν της εγκαταστάσεώς της), μέσω της δημιουργίας εικονιδίου ώστε να μπορεί να κληθεί χωρίς τη χρήση γραμμής εντολών. Για τους έμπειρους χρήστες Linux η περιγραφή της εγκατάστασης έγινε σχετικά εύκολα αποδεκτή. Γεγονός είναι πάντως ότι μπορεί η εγκατάσταση να απλουστευθεί με την ενσωμάτωση των δύο εφαρμογών και των απαιτούμενων βιβλιοθηκών σε ένα ενιαίο «πακέτο εγκατάστασης» (package).
- Η χρήση, κατά την εκτέλεση της εφαρμογής, φαίνεται να ξαφνιάζονται ευχάριστα από τον τρόπο με τον οποίο αυτή λειτουργεί. Γενικώς όμως δε φαίνεται να γίνεται συνειρμός μεταξύ του απεικονιζόμενου θέματος και του συνοδευτικού ήχου. Αυτό μάλλον οφείλεται στο ότι κατά τη διάρκεια της προβολής, η εφαρμογή δεν κάνει προφανές ποιο ακριβώς τμήμα της φωτογραφίας αποδίδεται ηχητικά κάθε χρονική στιγμή. Ένας άλλος λόγος σίγουρα είναι ότι δεν “αξιοποιείται ηχητικά” το σύνολο της φωτογραφίας αλλά ένα μόνο απόσπασμα. Βέβαια ίσως αυτό υποδηλώνει ότι για μια πιο αποτελεσματική υλοποίηση της συγκεκριμένης ιδέας απαιτείται στοχευμένη εκτενής μελέτη του ψυχοακουστικού παράγοντα.

7 ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Συμπεράσματα

Ο σχεδιασμός και η ανάπτυξη της εφαρμογής STGen πέρα από εκπαιδευτικό χαρακτήρα εμπειρείχε και το στοιχείο της πρόκλησης τόσο λόγω της μη-συμβατικότητας του θέματος όσο και της σχεδόν παντελούς έλλειψης υποδομής του συγγραφέα σε θέματα ανάλυσης και επεξεργασίας εικόνας και παραγωγής ήχου. Ωστόσο αποδείχτηκε εξαιρετικά πρόσφορο μέσο ώστε να εφαρμοστεί σημαντικό τμήμα των γνώσεων και δεξιοτήτων που αποκτήθηκαν τα τελευταία χρόνια στα πλαίσια της μεταπτυχιακής εκπαίδευσης αλλά και σπουδαίο πεδίο απόκτησης νέων.

Αξιολογώντας το τελικό αποτέλεσμα μπορούμε να πούμε τα παρακάτω:

- Ευκολία Χρήσης. Η εφαρμογή φαίνεται να κάνει προφανή τον τρόπο χρήσης της.
- Αποδοτικότητα Λειτουργίας. Η προέκταση της λειτουργίας της υπάρχουσας εφαρμογής με το στοιχείο της παραγωγής ηχητικής επένδυσης δεν εισάγει κάποια καθυστέρηση στο χρόνο απόκρισης.
- Ευκολία εκμάθησης. Η εφαρμογή δεν απαιτεί εκμάθηση πέρα από εκείνη που απαιτείται για την χρήση λογισμικού προβολής ροής εικόνων (δίχως την ηχητική επένδυση) για το μέσο χρήστη. Ωστόσο για τον χρήστη-ερευνητή απαιτείται εκμάθηση της χρήσης των επιμέρους εφαρμογών αν επιθυμεί να επέμβει τροποποιώντας τις αρχές διαχείρισης δεδομένων της εφαρμογής ή και τις αρχές λειτουργίας της.
- Αισθητικό αποτέλεσμα. Η ευρετική φύση των αλγορίθμων οι οποίοι χρησιμοποιήθηκαν για την παραγωγή των ήχων σίγουρα επιδέχονται βελτίωσης. Οποσδήποτε μπορούν να δίνουν ηχητικό αποτέλεσμα πιο συναφές με την προβαλλόμενη ροή εικόνων αλλά και να παράγουν πιο αρμονικό ηχητικό προϊόν. Ωστόσο δεδομένων των περιορισμών που τίθενται και από την εφαρμογή, το αποτέλεσμα φαίνεται να είναι αποδεκτό από τους χρήστες. Μια πιο «ασφαλής» προσέγγιση θα μπορούσε να περιορίζεται στην παραγωγή μόνον ήχων παρασκηνίου (π.χ ήχος ανέμου, βροχής, ή και ο ήχος απαλών κρουστών -ο οποίος ήδη χρησιμοποιείται και γίνεται ακουστός όταν η προβαλλόμενη εικόνα είναι μονοχρωματική). Σε αυτήν την περίπτωση όμως ο άμεσος συσχετισμός

ήχου και γεωμετρικών στοιχείων της εικόνας θα εξέλειπε. Ο λόγος (ο οποίος και έγινε αντιληπτός κατά την ανάπτυξη) είναι ότι, όπως είδαμε στο εδάφιο *‘Ζητήματα κατά την αξιοποίηση δεδομένων εικόνας’*, η πυκνότητα πληροφορίας που χρειάζεται ώστε να «αναπαρασταθεί ηχητικά» μια εικόνα είναι πολύ υψηλή και συνεπώς πυροδοτεί αλληπάλληλη ηχητική έξοδο. Όμως οι ήχοι παρασκηνίου και ιδίως οι «φυσικοί» ήχοι συνήθως διακρίνονται από ευδιάκριτη διάρκεια και μικρές αυξομειώσεις σε τονικότητα, ένταση και χροιά. Συνεπώς δεν είναι σε θέση να αποδώσουν τα γεωμετρικά στοιχεία της φωτογραφίας.

Προοπτικές Ανάπτυξης

Το STGen σε καμία περίπτωση δεν εξαντλεί τα περιθώρια ανάπτυξης που υπάρχουν σε αυτό το πεδίο εφαρμογών. Αποτελεί μόνο μια προσέγγιση σε ένα θέμα το οποίο μπορεί να οδηγήσει σε εφαρμογές ικανές να παράγουν σημαντικής αισθητικής ηχητικές επενδύσεις. Είναι ωστόσο μια ένδειξη του τι μπορεί να επιτευχθεί χρησιμοποιώντας εργαλεία ανάπτυξης και επαναχρησιμοποιώντας λογισμικό που διατίθεται δωρεάν ακόμη και σε περίπτωση που απουσιάζει αρχικά η υποδομή στο αντικείμενο της εφαρμογής.

Παρατίθενται σε αυτή τη θέση μερικές προτάσεις και σκέψεις για παραπέρα ανάπτυξη της εφαρμογής ή ακόμη και για εφαρμογές της αρχικής ιδέας σε άλλα συστήματα λογισμικού:

- Μια μικρή αλλά αρκετά σημαντική εξέλιξη της υπάρχουσας εφαρμογής θα μπορούσε να είναι μια μικρή αναπαράσταση/ένδειξη προς το χρήστη όσον αφορά το ποιο τμήμα της φωτογραφίας αξιοποιείται για την παραγωγή των ειδικών δεδομένων του ήχου.
- Η σε μεγαλύτερο βάθος μελέτη των αλγορίθμων όσον αφορά την ψυχοακουστική τους διάσταση θα μπορούσε να οδηγήσει σε ηχητικό αποτέλεσμα “κοντινότερο”

στις προβαλλόμενες εικόνες.

- Προς την κατεύθυνση αυτή, θα μπορούσε να υλοποιηθεί ένας “εκπαιδευόμενος” έξυπνος αλγόριθμος ο οποίος με αλληλεπίδραση με το χρήστη να βελτιώνεται όσον αφορά τις ηχητικές αποδόσεις κάθε προβαλλόμενης εικόνας. Ενώ μια εξελιγμένη έκδοση αυτής της ιδέας θα μπορούσε να αξιοποιεί ανίχνευση του σημείου εστίασης του χρήστη [53] ώστε να μπορεί να παραχθεί ηχητική επένδυση η οποία να λαμβάνει και αυτήν την παράμετρο υπόψιν.
- Μια απλούστερη προσέγγιση θα μπορούσε να συνίσταται από το να παρέχεται στο χρήστη η δυνατότητα να ελέγξει κάποιες τουλάχιστον παραμέτρους του ήχου.
- Εάν θεωρήσουμε ότι γίνεται εφικτή η συγκέντρωση της αξιολόγησης των χρηστών σχετικά με την καταλληλότητα ηχητικής επένδυση για τις εικόνες, θα είχε ενδιαφέρον να μελετηθεί η δυνατότητα δημιουργία βάσης δεδομένων προσβάσιμης μέσω δικτύου. Στη συνέχεια η εφαρμογή θα ήταν σε θέση να πραγματοποιεί πιο εύστοχες ηχητικές επιλογές.
- Τέλος μια άλλη ιδέα θα μπορούσε να είναι το να υλοποιηθεί η εφαρμογή ως πρόσθετο (plugin) σε υπάρχουσες εφαρμογές προβολής ροών στατικών εικόνων ή και εφαρμογές παρουσιάσεων.

Σίγουρα οι προοπτικές ανάπτυξης είναι μεγάλες και οι δυνατότητες εξέλιξης της ιδέας πίσω από την συγκεκριμένη εφαρμογή περιορίζονται μόνο από το βαθμό επινοητικότητας και το χρόνο ενασχόλησης του ερευνητή. Πραγματική επιτυχία της εφαρμογής πάντως θα ήταν το κέντρισμα του ενδιαφέροντος του απλού χρήστη ώστε να ασχοληθεί και ως ερευνητής με τον πραγματικά ανεξάντλητο κόσμο της εικόνας, του ήχου και της μουσικής.

8 ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Seeing, Hearing, and Smelling the world, a report from the Howard Hughes Medical Institute, 1995.

<http://www.hhmi.org/senses/senses.pdf>

[3] Εμμανουήλ Χελιδόνης, Σύγχρονη Ωτορινολαρυγγολογία, εκδ. Πασχαλίδη.

[4] Piano key frequencies - Wikipedia, the free encyclopedia,

http://en.wikipedia.org/wiki/Piano_key_frequencies

[5] Visible spectrum - Wikipedia, the free encyclopedia,

http://en.wikipedia.org/wiki/Visible_spectrum

[6] Auditory Perception and Sound as Event:

Theorising Sound Imagery in Psychology

<http://www.kent.ac.uk/arts/sound-journal/forrester001.html>

[7] Sound as Virtual Image, Bill Fontana

<http://www.resoundings.org/Pages/sound%20As%20Virtual%20Image.html>

[8] Silent film - Wikipedia, the free encyclopedia,

http://en.wikipedia.org/wiki/Silent_film

[9] First Retina Implant Surgery in the UK

<http://www.eye.ox.ac.uk/research/clinical-ophthalmology-research-group/press-releases/first-retina-implant-surgery-in-the-uk>

[10] William H. Dobbelle - Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/William_H._Dobbelle

[11] Bioinstrumentation - Artificial Vision History

<http://bme.sunysb.edu/labs/wlin/research/ArtificialVision.html>

[12] Augmented Reality for the Totally Blind

<http://www.seeingwithsound.com/>

[13] Visual resolution with The vOICe (sensory substitution for the blind)

<http://www.youtube.com/watch?v=GdOLDR6JHkk&fmt=18>

[14] Cortical Stimulation For The Evocation of Visual Perception

<http://www.scribd.com/doc/27146874/04-Wage>

[15] Neil Harbisson - Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Neil_Harbisson

[16] Γενικές Αρχές και Κλινική Πράξη της Ιατρικής Παθολογίας, Stanley Davidson et al.

Επιστημονικές Εκδόσεις Παρισιάνου, 19η Έκδοση

[17] Sonic Picture

<http://www.sonicpicture.de/>

[18] Music Visualization: Beautiful Tools to 'See' Sound

<http://www.visualcomplexity.com/vc/blog/?p=811>

[19] Narratives, Matthias Dittrich

<http://www.matthiasdittrich.com/>

[20] MuSA.RT - music on the spiral array . real-time

<http://www-bcf.usc.edu/~mucoaco/MuSA.RT/>

[21] Spectral Music - Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Spectral_music

[22] Lempel-Ziv Data Compression, 1977

http://www.ieeeahn.org/wiki/index.php/Milestones:Lempel-Ziv_Data_Compression_Algorithm,_1977

[23] D.A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes", Proceedings of the I.R.E., September 1952

[24] Scalable Vector Graphics (SVG) 1.2 W3C Working Draft. World Wide Web Consortium. 15 November 2002. 29 August 2010.

[25] PDF Reference and Adobe Extensions to the PDF Specification

http://www.adobe.com/devnet/pdf/pdf_reference.html

[26] John Canny. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8(6):679–698, Nov. 1986.

[27] Whitten, Jeffrey L.; Lonnie D. Bentley, Kevin C. Dittman. *Systems Analysis and*

Design Methods. 6th edition. Nov. 2003.

[28] Principles behind the Agile Manifesto

<http://www.agilemanifesto.org/>

[29] K Petersen's doctoral research in Sweden [Implementing Lean and Agile Software Development in Industry](#)

[30] Feller, J., Fitzgerald, B. & Hissam, S. A. (eds.), (2005): Perspectives on Free and Open Source Software, MIT Press.

[31] Benkler, Y. (December 2002). "[Coase's Penguin, or, Linux and The Nature of the Firm](#)" (PDF). *Yale Law Journal* **112** (3): 367(78).

[32] GNU General Public License

<http://www.gnu.org/licenses/gpl.html>

[33] Patterns of Open Innovation in Open Source Software

http://openinnovation.berkeley.edu/ranp_chapters/05.pdf

[34] Shotwell Photo Organizer

<http://www.yorba.org/projects/shotwell/>

[35] Csound - Wikipedia, the free encyclopedia,

<http://en.wikipedia.org/wiki/Csound>

[36] Eclipse

<http://www.eclipse.org/>

[37] QuteCsound

<http://qutecsound.sourceforge.net/>

[38] Wireshark

<http://www.wireshark.org/>

[39] Introduction to OSC

<http://opensoundcontrol.org/introduction-osc>

[40] Vala – Compiler for the GObject type system

<https://wiki.gnome.org/Vala>

[41] GDB: The GNU Project Debugger

<http://www.gnu.org/software/gdb/>

[42] liblo: Lightweight OSC implementation

<http://liblo.sourceforge.net/>

[43] Binding Libraries with VAPI Files

https://wiki.gnome.org/Vala/Tutorial#Binding_Libraries_with_VAPI_Files

[44] Shotwell Architecture Overview

<http://redmine.yorba.org/projects/shotwell/wiki/ShotwellArchitectureOverview>

[45] The Canonical Csound Reference Manual

<http://www.csounds.com/manual/html/index.html>

[46] fork(2) - Linux man page

<http://linux.die.net/man/2/fork>

[47] Unified File Format for Orchestras and Scores

<http://csounds.com/manual/html/CommandUnifile.html>

[48] User Datagram Protocol

<http://www.ietf.org/rfc/rfc768.txt>

[49] Transmission Control Protocol

<http://www.ietf.org/rfc/rfc793.txt>

[50] How many pixels make an image?

http://journals.cambridge.org/download.php?file=%2FVNS%2FVNS26_01%2FS0952523808080930a.pdf&code=b2a050f0e040b0159a5cdd4ecca8b84c

[51] Meet Ubuntu

<http://www.ubuntu.com/desktop>

[52] Krug, Steve – Don't Make Me Think – A Common Sense Approach to Web Usability, 2nd Edition, 2006

[53] Eye tracking - Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Eye_tracking

9 ΠΑΡΑΡΤΗΜΑ Α

Στο παράρτημα αυτό αναφέρονται οι οδηγίες για τα απαιτούμενα βήματα ώστε να παραμετροποιηθεί και να παραχθεί το εκτελέσιμο.

Τα προ-απαιτούμενα για να παραχθεί το εκτελέσιμο είναι κάποιο περιβάλλον Linux (οι διαδικασίες δοκιμάστηκαν στη διανομή Ubuntu 11.10 αλλά η εκτέλεση είναι εφικτή σε όλες τις διανομές στις οποίες είναι δυνατή η εγκατάσταση του Shotwell και Csound).

-Αποθηκεύουμε τοπικά στον σκληρό δίσκο το αρχείο από το σύνδεσμο <https://rapidshare.com/files/2118872257/stgen.tgz>

-Μέσω τερματικού, βρίσκουμε την τοποθεσία όπου αποθηκεύτηκε το αρχείο stgen.tgz ή εναλλακτικά μέσω του φυλλομετρητή αρχείων

- Αποσυμπιέζουμε το αρχείο είτε με την εντολή:

```
tar -xvzf stgen.tgz
```

είτε με τη βοήθεια κάποιου άλλου λογισμικού το οποίο να υποστηρίζει αποσυμπίεση αρχείων .tgz

- Για την εγκατάσταση των απαιτούμενων βιβλιοθηκών για το Shotwell χρησιμοποιούμε την παρακάτω εντολή (σε διανομές διαφορετικές του Ubuntu θα πρέπει να χρησιμοποιηθεί η αντίστοιχη εντολή εγκατάστασης):

```
sudo apt-get install desktop-file-utils libgconf2-dev libgee-dev libgexiv2-dev libglib2.0-dev libgstreamer1.0-dev libgstreamer-plugins-base1.0-dev libgtk-3-dev libgudev-1.0-dev libexif-dev libgphoto2-2-dev libraw-dev librest-dev libsoup2.4-dev libxml2-dev libsqlite3-dev m4 valac-0.18 libjson-glib-dev
```

- Εντός του φακέλλου “*shotwell-0.12.3*” δίνουμε την εντολή:

```
./configure
```

- Για την εγκατάσταση του Csound χρειάζονται τα παρακάτω:

Εγκατάσταση του libsndfile έκδοση 1.0.13 ή μεταγενέστερη από το www.mega-

nerd.com/libsndfile [<http://www.mega-nerd.com/libsndfile>].

Εγκατάσταση της Python από www.python.org [<http://www.python.org>]. (η εγκατάσταση απαιτείται ώστε να μπορεί να λειτουργήσει ο μηχανισμός “χτισίματος” SCons που χρησιμοποιεί το Csound)

Εγκατάσταση του SCons από www.scons.org [<http://www.scons.org>].

Εγκατάσταση της βιβλιοθήκης PortAudio (για ήχο πραγματικού χρόνου) από www.portaudio.com/usingcv.html

[<http://www.portaudio.com/usingcv.html>]. Σημείωση: Η έκδοση 18 της βιβλιοθήκης PortAudio δεν λειτουργεί σωστά.

- Εντός του φακέλλου “csound5” δίνουμε την εντολή:

```
./configure
```

- Ρύθμιση της παραμέτρου περιβάλλοντος PKG_CONFIG_PATH με την παρακάτω εντολή:

```
export PKG_CONFIG_PATH=/usr/local/lib/pkgconfig
```

(αυτό το βήμα χρειάζεται ώστε να γίνει “ορατή” στο Shotwell η βιβλιοθήκη liblo)

Για το χτίσιμο της -ενιαίας- εφαρμογής αρκούν οι παρακάτω εντολές:

```
export OPCODEDIR64=/usr/lib/csound
```

```
scons buildCsound5GUI=1 buildPythonOpCodes=1 useOSC=1 builBeats=1 install=1 lib64=1 usePortAudio=1
```

(αλλά αυτό εξαρτάται και από το εάν χρησιμοποιείται η lib64 ή όχι)

Για να εκτελέσουμε την εφαρμογή χρησιμοποιούμε (εντός του φακέλλου csound5) την παρακάτω εντολή:

```
./csound -+rtaudio=alsa -o dac -i adc examples/STGEN.csd (1)
```

(εαν επιθυμούμε καταγραφή συμβάντων (log files) από το Shotwell, πριν από την παραπάνω εντολή, δίνουμε:

```
export SHOTWELL_LOG=1
```

Σημείωση: Το αρχείο καταγραφής συμβάντων βρίσκεται στο φάκελλο `~/.cache/shotwell/shotwell.log`)

Με την εντολή (1) εμφανίζονται δύο παράθυρα (Csound και Shotwell)

Η εισαγωγή φωτογραφιών στο Shotwell μπορεί να γίνει με οποιονδήποτε από της μεθόδους που περιγράφονται εδώ:

<http://www.yorba.org/shotwell/help/>

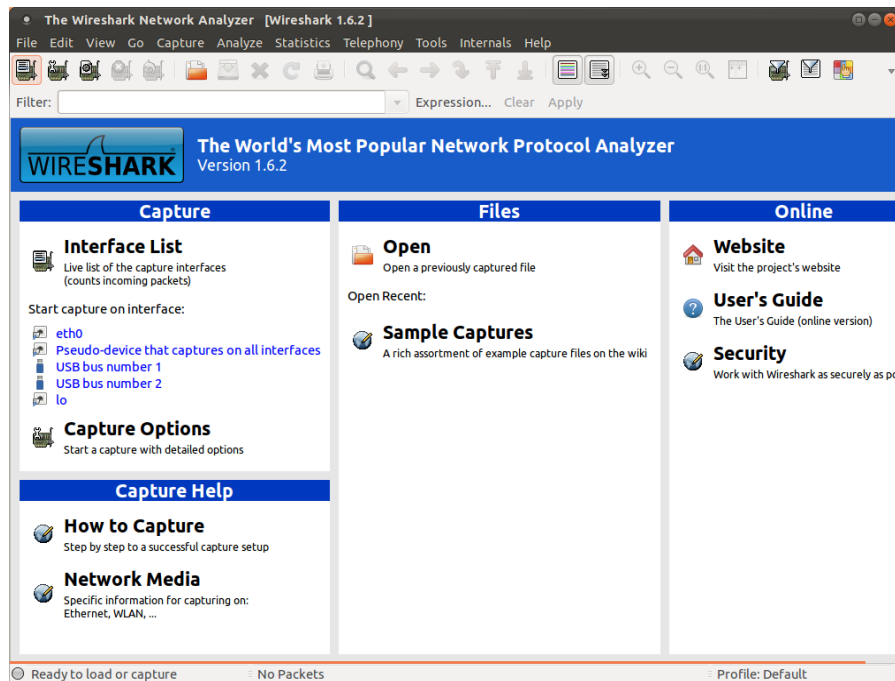
Για την προβολή της ροής εικόνων με αυτόματη ηχητική επένδυση επιλέγουμε κάποια από τις διαθέσιμες βιβλιοθήκες (προαιρετικά μπορούμε να επιλέξουμε και από ποια φωτογραφία επιθυμούμε να ξεκινήσει η ροή), και μέσα από τη γραμμή εργαλείων του Shotwell χρησιμοποιούμε το πλήκτρο View και από το αναδυόμενο μενού επιλέγουμε Slideshow with Sound

10 ΠΑΡΑΡΤΗΜΑ Β

Σε αυτό το παράρτημα γίνεται μια σύντομη αναφορά στον τρόπο με τον οποίο αξιοποιήθηκε το λογισμικό ανάλυσης πακέτων δεδομένων δικτύων υπολογιστών Wireshark[38] στα πλαίσια των αναγκών ανάπτυξης του συστήματος λογισμικού της εργασίας και η σχέση της χρήσης του με τα δεδομένα προφίλ που παράγονται τελικά από μιαν εικόνα σύμφωνα με αυτά που περιγράψαμε στο κεφάλαιο 3, Αλγόριθμος ανίχνευσης ακμών Canny.

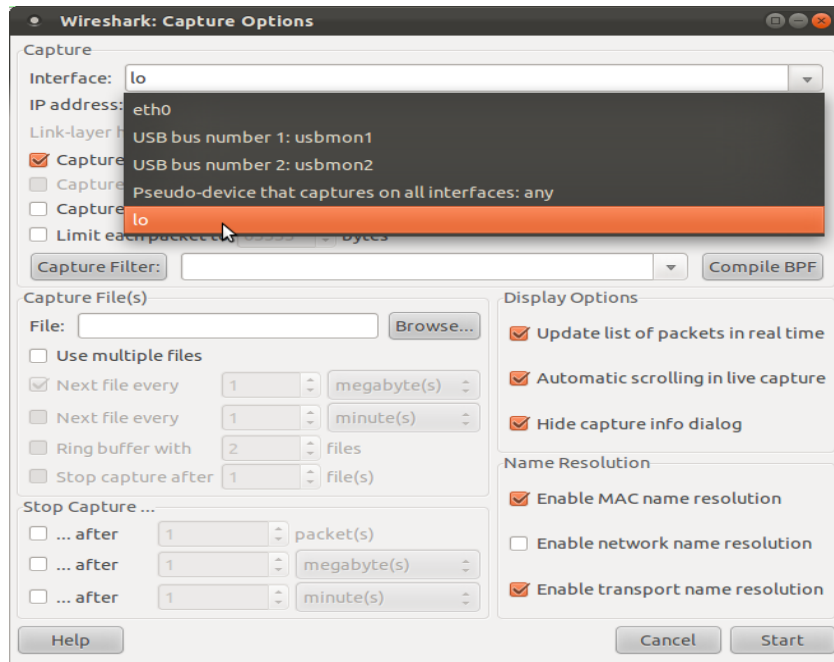
Το λογισμικό έχει τη δυνατότητα να ανιχνεύει πακέτα μιας πληθώρας πρωτοκόλλων δικτύων. Μπορούμε να το εκτελέσουμε με διπλό κλικ ή και μέσω του τερματικού.

Όταν εκτελείται το πρόγραμμα, εμφανίζεται το γραφικό περιβάλλον που φαίνεται παρακάτω:



Εικόνα 233 Η οθόνη υποδοχής του Wireshark

Όπου μας γνωστοποιείται ότι το λογισμικό είναι έτοιμο να φορτώσει αρχείο που να περιέχει πακέτα δεδομένων ή να ξεκινήσει τη σύλληψη νέων πακέτων δεδομένων. Μας παρέχει, μεταξύ άλλων, απευθείας πρόσβαση -μέσω συνδέσμων- σε ορισμένες λειτουργίες του λογισμικού (λ.χ. “Start capture on interface: eth0” κλπ.). Το πρωτόκολλο OSC δεν περιλαμβάνεται στη λίστα των γνωστών για το Wireshark πρωτοκόλλων ωστόσο μπορούμε να απομονώσουμε τα μηνύματα/πακέτα που μας ενδιαφέρουν μιας και οι υποδοχές (sockets) που χρησιμοποιούνται κατά την εκτέλεση της εφαρμογής μας είναι γνωστές (είναι προκαθορισμένες στον κώδικα) και το λογισμικό μας παρέχει τη δυνατότητα να φιλτράρουμε την πληροφορία με βάση τη θύρα προορισμού (destination port). Αυτό μπορούμε να το πετύχουμε συμπληρώνοντας στο πεδίο 'Filter' τις θύρες για τις οποίες ενδιαφερόμαστε (`udp.dstport == 7613 || udp.dstport == 7614`). Εναλλακτικά μπορούμε να επιλέξουμε από τις επιλογές ανίχνευσης πακέτων την ανίχνευση μόνο της διεπαφής lo όπως φαίνεται παρακάτω:



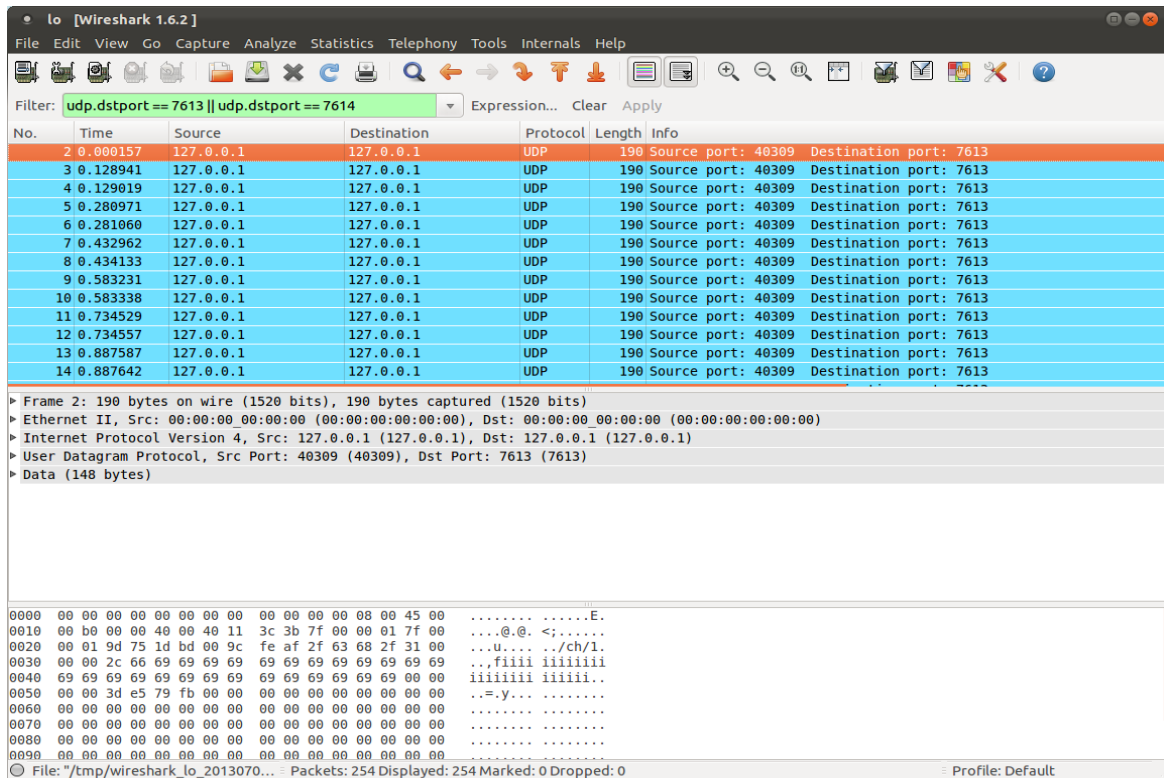
Εικόνα 244 Επιλογή διεπαφής για την ανίχνευση των πακέτων δεδομένων

Σημειώνεται εδώ ότι για να συμπεριλαμβάνεται από το Wireshark η διεπαφή lo μεταξύ των διαθέσιμων επιλογών, η εφαρμογή θα πρέπει να έχει εκτελεστεί με δικαιώματα υπερ-χρήστη (superuser). Για λειτουργικό σύστημα Linux, αυτό πραγματοποιείται με την εντολή sudo:

```
sudo /usr/bin/Wireshark
```

Στη συνέχεια, είτε μέσω της συντόμευσης στην επιφάνεια εργασίας είτε από ένα τερματικό, εκκινούμε την εφαρμογή μας και εκτελούμε τις απαραίτητες ενέργειες ώστε να εκκινήσουμε την προβολή εικόνων με αυτόματη ηχητική επένδυση (βλ. Παράρτημα Α για οδηγίες).

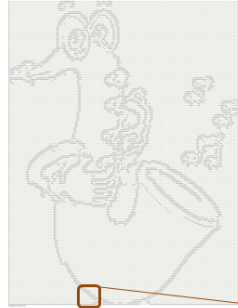
Στην παρακάτω εικόνα απεικονίζονται τα πακέτα που αντιστοιχούν σε ροή στατικών εικόνων διάρκειας μερικών δευτερολέπτων.



Εικόνα 255 Συλλογή πακέτων δεδομένων

Στο πακέτο με αριθμό 2 μπορούμε να διακρίνουμε τους παρακάτω τύπους δεδομένων:

- Χαρακτήρες (π.χ. '/ch/I'): υποδηλώνουν στον παραλήπτη ότι το μήνυμα προορίζεται για το Csound και ταυτόχρονα μπορεί να προσδιορίζουν και τον τύπο του μηνύματος.
- Αλφαριθμητικούς χαρακτήρες (π.χ. 'fii iii...'): προσδιορίζουν τον τύπο των δεδομένων που ακολουθούν. Το 'f' αντιπροσωπεύει μια τιμή κινητής υποδιαστολής (float) ενώ το 'i' αντιπροσωπεύει έναν ακέραιο αριθμό (integer)
- Αριθμητικές τιμές (π.χ. '00 00 00...'): οι τιμές των δεδομένων ήχου. Στο μήνυμα που απεικονίζεται, οι μηδενικές τιμές αντιστοιχούν σε προφίλ/κάθετη τομή λευκής φωτογραφίας. Τέτοιες φωτογραφίες χρησιμοποιήθηκαν εκτενώς για λόγους ελέγχου ορθής λειτουργίας της εφαρμογής. Αξίζει να αναφέρουμε εδώ ότι το λευκό σε όρους τιμών RGB (κόκκινου-πράσινου-μπλε) και σε format εικόνας όπου τα επίπεδα παρουσίας κάθε χρώματος αντιπροσωπεύονται από 8 bits ανά χρώμα, αντιστοιχεί σε R=255, G=255, B=255. Συνεπώς θα περίμενε κανείς και η συνισταμένη τιμή που αντιστοιχεί σε ένα λευκό εικονοστοιχείο να είναι 255 και η μηδενική τιμή να αντιστοιχεί σε μαύρο



```
0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 255 255 255 255
```

Εικόνα 277 Εξαγωγή προφίλ ήχου