

Ψηφιακός ήχος και ...μύθος

Αλέξανδρος Μούζας

Στο παρόν άρθρο γίνεται αναφορά σε βασικές έννοιες που σχετίζονται με τον ψηφιακό ήχο και ελπίζουμε να βοηθηθεί ο αναγνώστης στην κατανόηση προβλημάτων και την εξεύρεση αποτελεσματικών λύσεων.

Καταρχήν, οτιδήποτε υπάρχει, αποθηκεύεται ή εμφανίζεται στον υπολογιστή είναι αναπαράσταση ψηφιακής πληροφορίας. Δεδομένα που εισέρχονται στις εισόδους, αποθηκεύονται ή εξέρχονται είναι απλώς αριθμοί, και μάλιστα της μορφής 0 και 1 (bits). Ο τεράστιος αριθμός των bits (που ανά 8 σχηματίζουν ένα byte) είναι μια σειρά από data που αποκτούν νόημα και γίνονται πληροφορία (information), μέσω του λειτουργικού συστήματος και των εξειδικευμένων προγραμμάτων.

Να θυμάστε: Ο υπολογιστής δεν μπορεί να αναγνωρίσει τίποτε άλλο πέραν του 0 και 1.

Ο ήχος μετατρέπεται από αναλογικός σε ψηφιακό (A/D conversion), και αφού αποθηκευτεί (digital storage), μπορεί να επεξεργαστεί (digital processing) και να μεταφερθεί ή να αναπαραχθεί (D/A conversion).

Σε όλη αυτή τη ψηφιακή διαδικασία είναι λάθος να θεωρήσουμε ότι δεν υπάρχουν προβλήματα και απώλειες.

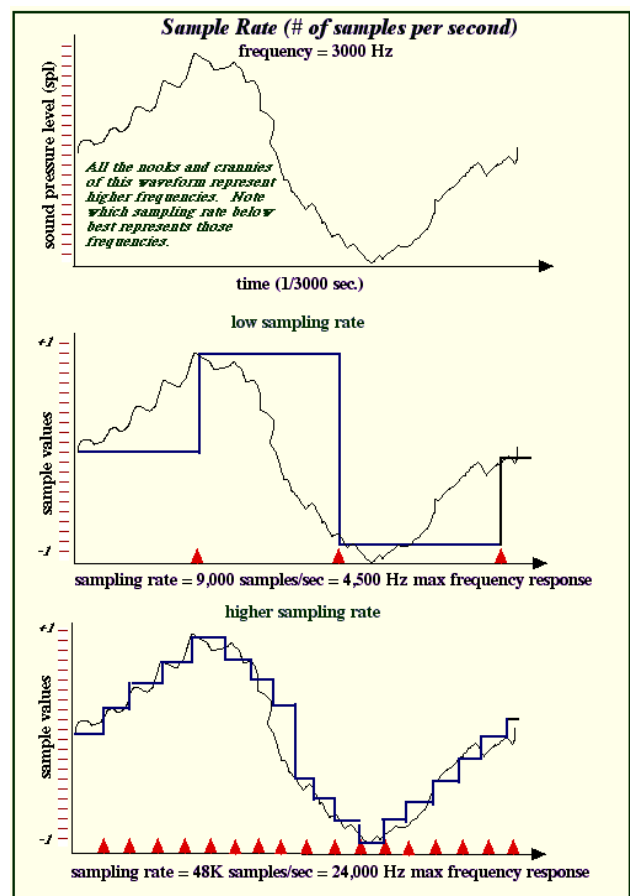
Βασική έννοια του ψηφιακού ήχου είναι η δειγματοληψία (sampling).

Ο αριθμός των σημείων μιας συνεχούς καμπύλης, που αφορά στην μεταβολή της πίεσης του αέρα ή στη μεταβολή του τάσης του ρεύματος ενός αναλογικού σήματος (κυματομορφή), μετράται από το σύστημα σε τακτά χρονικά διαστήματα και αποτυπώνεται με μορφή αριθμών. Οι μετρήσεις αυτές βαλμένες στη σειρά επανασηματίζουν την κυματομορφή του αρχικού ήχου. Ένα άπειρο μέγεθος, δηλαδή, αποτυπώνεται με πεπερασμένο αριθμό μετρήσεων. Η καμπύλη των άπειρων σημείων της κυματομορφής αναπαράγεται από σειρά διαδοχικών σημείων, που όσο πιο πυκνά είναι τόσο πλησιάζουν την συνεχή γραμμή (χωρίς όμως να την φτάνουν ποτέ).

Η συχνότητα δειγματοληψίας (Sampling

Rate ή SR) είναι ο αριθμός των δειγμάτων ή μετρήσεων ανά δευτερόλεπτο. Στο CD η συχνότητα αυτή είναι 44100 Hz. Το δείγμα δηλαδή χωρίζεται και "φωτογραφίζεται" 44100 το δευτερόλεπτο. Όμως, προσοχή: οτιδήποτε μεσολαβεί μεταξύ δύο διαδοχικών μετρήσεων, αγνοείται και δεν καταγράφεται. Ο ψηφιακός ήχος είναι γενικά δειγματοληπτική διαδικασία.

Φανταστείτε πως λειτουργεί η κινηματογραφική



κή κάμερα. Καταγράφει στιγμιαίες εικόνες που, στην αναπαραγωγή, καθώς περνούν αστραπιαία μπροστά από τα μάτια μας, δίνουν την εντύπωση της αρχικής κινούμενης εικόνας.

Σύμφωνα με τον νόμο του Nyquist, το SR πρέπει να είναι τουλάχιστον διπλάσιο της μέγιστης συχνότητας. Καθώς η μέγιστη συχνότητα του ακουστικού μας φάσματος είναι 20KHz, το SR πρέπει να είναι τουλάχιστον 40000Hz. Στο CD για τεχνικούς λόγους χρησιμοποιούμε την τιμή 44100. Οποιοσδήποτε συχνότητες πέραν των 22000Hz, που ενώ υπάρχουν στη φύση, αλλά δεν είμαστε ικανοί να τις ακούσουμε, δεν καταγράφονται. Το ψηφιακό σύστημα λειτουργεί σαν φίλτρο που αποκλείει τις συχνότητες εκτός των καθορισμένων ορίων.

(Αν η γάτα σας ενοχλείται από υπερήχους, ηχογραφήστε το ηχητικό περιβάλλον σε CD και βάλτε να ακούσει την ηχογράφιση με ακουστικά. Είναι σίγουρο ότι θα νιώσει πολύ καλύτερα.)

Για να αντιληφθείτε την έννοια του SR, παρατηρήστε από πολύ κοντά μια φωτογραφία σε ένα έντυπο. Θα παρατηρήσετε ότι η εικόνα συντίθεται από μικροσκοπικές κουκκίδες (pixels), που όσο πιο πυκνές είναι τόσο η φωτογραφία απεικονίζει πιο πιστά το αυθεντικό αντικείμενο. Ο αριθμός τους και η πυκνότητά τους ανά ίντσα είναι το αντίστοιχο του SR.

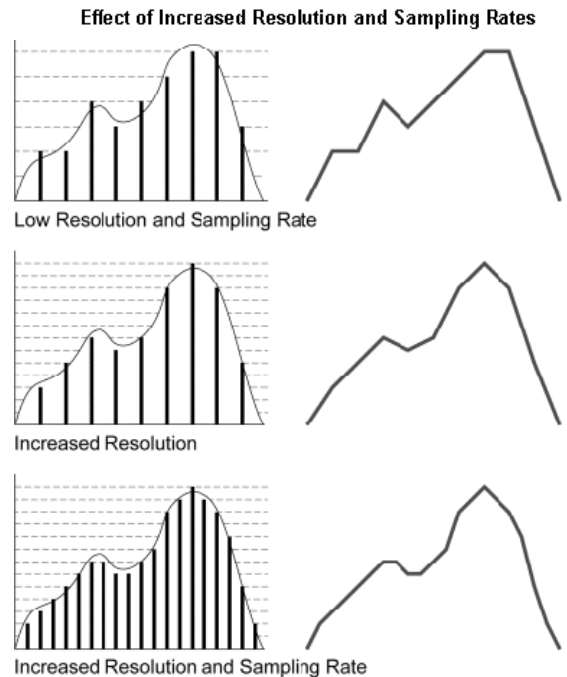
Ο μεγαλύτερος ρυθμός δειγματοληψίας επιτρέπει στο ψηφιακό σύστημα να κάνει μετρήσεις σε μικρότερα διαστήματα και, καθώς ο ήχος είναι συνάρτηση της συχνότητας, μπορεί να καταγράφει και υψηλότερες συχνότητες. Οι τιμές ποικίλουν: 44.100Hz, 48.050Hz, 96KHz, 192KHz.

Αν η δειγματοληψία είναι χαμηλής συχνότητας, η παραγόμενη καμπύλη δεν ακολουθεί πιστά το αρχική κυματομορφή, δημιουργείται σφάλμα και παραμόρφωση αλλά και η αρμονική παραμόρφωση aliasing, που υπεισέρχεται στο ακουστικό φάσμα επιτείνει το αποτέλεσμα.

| rate | use(s) |
|-------|--|
| 32K | older DATs, voice quality |
| 44.1K | CD, DAT, digital recording software/hardware |
| 48K | DAT, digital recording software/hardware |
| 96K | digital recording software/hardware |
| 192K | digital recording software/hardware |

Δεύτερη παράμετρος που μας ενδιαφέρει είναι η **ανάλυση (Resolution)**.

Για κάθε δειγματοληψία ο υπολογιστής αποθηκεύει την μέτρηση του σε ψηφιακά δεδομένα. Στον κάθετο άξονα του σχήματος μας μετράμε το ύψος της κυματομορφής που αντιστοιχεί στην ένταση κάθε στιγμής



(amplitude).

Μεταξύ μέγιστης και ελάχιστης τιμής υπάρχουν άπειρες διαβαθμίσεις, που, καθώς, ο υπολογιστής δεν μπορεί να αναγνωρίσει, χωρίζει την περιοχή σε ίσα μέρη και καταγράφει πεπερασμένο πλήθος μετρήσεων έντασης. Αν για κάθε δείγμα, που η πυκνότητα τους ορίζεται από την συχνότητα δειγματοληψίας, ο κατασκευαστής χρησιμοποιεί 1 byte, τότε οι δυνατές τιμές που μπορεί να εκφράσει είναι 256 (2⁸). Ένα 8μπιτο (1byte) ψηφιακό σύστημα, δηλαδή, μπορεί να παράγει 256 μόνο διαφορετικές βαθμίδες έντασης. Το σφάλμα (σφάλμα κβαντοποίησης) που δημιουργείται είναι αρκετά μεγάλο και η κυματομορφή που παράγεται απέχει αρκετά από την αρχική.

Στο CD για κάθε μέτρηση αποθηκεύονται 2 byte (16bit) και μπορούν να εκφράσουν 256*256=65.536 διαφορετικές τιμές έντασης.

Παρόλο που η τιμή των 16 bit θεωρείται ικανή για την πιστή αναπαραγωγή του ήχου, εξακολουθεί το σύστημα να χρησιμοποιεί πεπερασμένο αριθμό για ένα άπειρο μέγεθος. Στην επεξεργασία, όμως, για να ελαχιστοποιήσουμε τον κβαντικό θόρυβο, απαιτείται μεγαλύτερη ακρίβεια, και κρίνεται απαραίτητη η χρήση της μεγαλύτερης δυνατής ανάλυσης σε κάθε στάδιο της διαδικασίας: Οποιαδήποτε είδους επεξεργασία, ακόμη και απλή αλλαγή στάθμης, προκαλεί ανεπιθύμη και μόνιμη ζημιά στο ψηφιακό σήμα.

Προσέξτε το παράδειγμα:

Ας υποθέσουμε ότι σε κάποια χρονική στιγμή η ένταση της κυματομορφής έχει την τιμή 30.001 (στο εύρος 0- 65.536). Ας υποθέσουμε επίσης ότι με μια απλή κίνηση του fader του sequencer, θέλουμε να φέρουμε την ένταση στο ήμισυ (κάτι τέτοιο δεν στέκει στην φυσική, καθώς τα μεγέθη είναι λογαριθμικά, αλλά το χρησιμοποιούμε χάριν του παραδείγματος). Η τιμή της στιγμιαίας έντασης θα γίνει τότε 15.000,5. Καθώς, δεκαδικά ψηφία δεν μπορούν να εκφραστούν, το σύστημά πρέπει να επιλέξει μεταξύ των τιμών 15.000 και 15.001. Ήδη δηλαδή δημιουργήθηκε ένα σφάλμα, και μάλιστα τάξεως $1/65.536$. Αν ένας ήχος επιδέχεται συνεχούς επεξεργασίας (εφέ, μίξη, mastering, κλπ) καταλαβαίνετε ότι το σφάλμα συνεχώς αυξάνεται. Γι αυτό το λόγο, σε επαγγελματικό επίπεδο, χρησιμοποιείτε ανάλυση 24bit (3 byte) που δίνει $256*256*256=16.777.216$ διαφορετικές τιμές έντασης. Ξαναγυρνώντας στο παράδειγμα μας, μια τιμή έντασης πχ 10.000.121 γίνεται τώρα είτε 5.000.061 ή 5.000.062. Σφάλμα φυσικά, αλλά τάξεως $1/16.777.216$, που θεωρείται αμελητέο και ελάχιστα επηρεάζει το ήχο. Στο τέλος της διαδικασίας, μετατρέπουμε το σήμα σε 16 bit, γιατί μόνο έτσι μπορεί να αναπαραχθεί το CD-Player.

Η αντιστοιχία που υπάρχει στην εικόνα αφορά στον αριθμό των διαφορετικών χρωμάτων που μπορεί να απεικονίσει ένα σύστημα. Από τη παλέτα άπειρων χρωματικών διαβαθμίσεων το ψηφιακό σύστημα, μπορεί να εκφράσει μόνο έναν συγκεκριμένο αριθμό. (256 χρώματα/8bit, χιλιάδες χρώματα/16bit ή εκατομμύρια χρώματα/24bit) Αφορά δε κάθε μία από τις κουκκίδες (pixels) που συνθέτουν την ψηφιακή εικόνα.

Δοκιμάστε στη οθόνη του υπολογιστή σας τις διαφορετικές ρυθμίσεις και τις φωτογραφίες σας σε διαφορετική ανάλυση και ποιότητα.

Τύπος

Ένας χρήσιμος τύπος για να υπολογίζεται ποσότητες πληροφορίας για εγγραφή και αποθήκευση ψηφιακού ήχου.

Συχν.Δειγματοληψίας * Ανάλυση * χρόνο * κανάλια = Ποσότητα πληροφορίας (byte)

ΠΧ. Σε ποιότητα CD ένα λεπτό μουσικής καταλαμβάνει 10MB περίπου.

$$44100 * 2 \text{ byte} (16\text{bit}) * 60 \text{ sec} * 2 \text{ (stereo)} = 10\text{MB}$$

Αντικαθιστώντας τις τιμές μπορείτε να προσαρμόσετε τον τύπο σε κάθε προδιαγραφή.

ΠΧ. Θέλουμε να αποθηκεύσουμε 16 κανάλια πολυκάναλης ηχογράφησης, με ανάλυση 24bit και δειγματοληψία 96KHz, συνολικής διάρκειας 3 λεπτών:

$$96.000 \text{ HZ} * 3 \text{ byte} (24\text{bit}) * 180\text{sec} * 16 \text{ tracks} = 767 \text{ MB}$$

Η μεγάλη δειγματοληψία και ανάλυση απαιτούν από τα συστήματα μεγαλύτερη μνήμη και επεξεργαστική ισχύ, ενώ επιβάλλουν επίπονη προσπάθεια και υψηλές ταχύτητες στους σκληρούς δίσκους.

| word size | power of 2 | # of values |
|------------------|------------|---------------|
| 8-bit (1 byte) | 2^8 | 256 |
| 16-bit (2 bytes) | 2^{16} | 65,536 |
| 24-bit (3 bytes) | 2^{24} | 16,777,216 |
| 32-bit (4 bytes) | 2^{32} | 4,294,967,296 |

Jitter

Η ψηφιοποίηση στηρίζεται στην έννοια της σταθερής δειγματοληψίας και απαιτεί, από τον ταλαντωτή της κάρτας ήχου, σταθερό ρολόι. Αν το ρολόι (digital clock), που παράγει περιοδικά παλμούς σε σειρές 0 και 1, είναι ασταθές, κατά την διάρκεια της μετατροπής οι διαφορετικές τιμές έντασης καταγράφονται σε ακανόνιστες χρονικά διαστήματα. Μια μικρή μετατόπιση του χρόνου λήψης προκαλεί παραγωγή λάθος λέξης, που σημαίνει παραμόρφωση του τελικού αποτελέσματος, θολή εικόνα, απώλεια βάθους, στερεοφωνικής εικόνας και χώρου. Το σφάλμα ονομάζεται jitter και δημιουργείται από την κακή ποιότητα του ψηφιακού ταλαντωτή, μπορεί δε να εμφανιστεί κάθε φορά που υπάρχει μετατροπή (sampling jitter) αλλά και επικοινωνία μεταξύ ψηφιακών συσκευών (interface jitter). Το παράδοξο είναι ότι όσο μεγαλώνει η συχνότητα δειγματοληψίας τόσο μεγαλώνει και το λάθος. Η λύση έγκειται στην αγορά ποιοτικών καρτών ήχου, στον συγχρονισμό της συσκευής μας με το εισερχόμενο ψηφιακό σήμα μέσω word clock ή στην περίπτωση των στούντιο, στην χρήση ειδικών συσκευών που παρέχουν master clock σε όλο τον ψηφιακό εξοπλισμό. (Το πρόβλημα δεν αφορά τα αναλογικά μηχανήματα και το word clock δεν έχει σχέση με τους κώδικες SMPTE /Midi Time Code)

Η χαμηλή στάθμη εγγραφής είναι ένα άλλο πρόβλημα που δεν επιτρέπει στο σύστημα να εκμεταλλευτεί όλες του τις δυνατότητες. Ένας μετατροπέας 16bit μπορεί να λειτουργεί ουσιαστικά μόνο με τα 12, 10 ή και λιγότερα bits προσθέτοντας στο σήμα θόρυβο και παραμόρφωση.

Οι κάρτες ήχου που κυκλοφορούν στην αγορά προσφέρουν όλες σχεδόν τις δυνατότητες που αναφέρουμε παραπάνω. Όμως συνοψίζοντας, ο κακός ψηφιακός ήχος προέρχεται από κακή κατασκευή, προβληματικά φίλτρα, χαμηλή δειγματοληψία, χαμηλή ανάλυση, ακατάλληλο dither, ασταθές ψηφιακό ρολόι. Δεν είναι τυχαίο ότι οι τιμές των AD/DA converters ποικίλουν από 100 έως 10.000 ευρώ.

Στο στούντιο πάντα ξεκινάμε με την μεγαλύτερη δυνατή ανάλυση και την διατηρούμε όσο περισσότερο μπορούμε μέσα στην διαδικασία της μίξης ή του mastering. Προτείνεται η ηχογράφηση να γίνεται σε 24bit, το τελικό stereo αρχείο σε 24bit και μόνο μετά το mastering να το κατεβάζουμε στα 16bit. Χρησιμοποιώντας μεγάλη δειγματοληψία χρειαζόμαστε λιγότερα ακριβά και ακριβή αντί-aliasing φίλτρα. Γίνεται μεγάλη συζήτηση για την διαφορά στην ποιότητα του ήχου στη μεγάλη δειγματοληψία (96Khz, 192Khz) και αν θα πρέπει να καταγράψουμε συχνότητες πέραν του ακουστικού μας φάσματος. Οι γνώμες δίστανται και πολλοί θεωρούν ότι οι πολύ υψηλές συχνότητες των οργάνων επηρεάζουν θετικά το ακουστικό αποτέλεσμα. Μερικοί μάλιστα πιστεύουν ότι η διαφορά οφείλεται στις νότες διαφοράς που υπεισέρχονται στο ακουστικό φάσμα, και οι οποίες εξαφανίζονται καθώς οι γενεσιουργές συχνότητες αποκλείονται (φιλτράρονται) στην εγγραφή.

Τι λένε οι ειδικοί:

Αυτή τη στιγμή, το καλύτερο μέσο καταγραφής της τελικής μίξης θεωρείται η ταινία μαγνητοφώνου 1/2 ίντσας χωρίς DOLBY σε ταχύτητα 30ips.

...και ένα παράδοξο

Ξοδεύουμε χρόνο και χρήμα στην έρευνα και εξοπλισμό για να προσεγγίσουμε τον τέλειο ήχο, για να τον συμπιέσουμε τελικά σε mp3. Μήπως ο μέσος ακροατής πιστεύει ότι η ποιότητα που του προσφέρει το CD είναι ήδη παραπάνω από ότι πραγματικά χρειάζεται?

Το παρόν άρθρο κυκλοφόρησε στο περιοδικό «Πολύτονον» τον Αύγουστο 2006