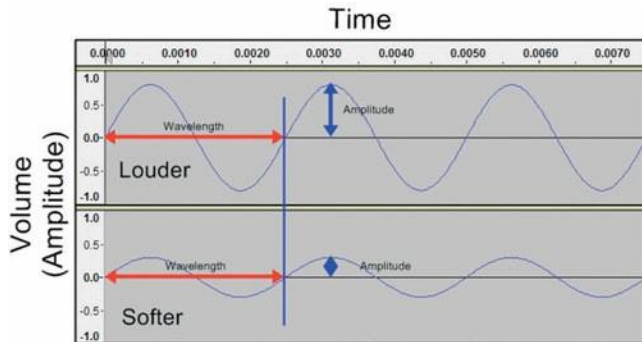


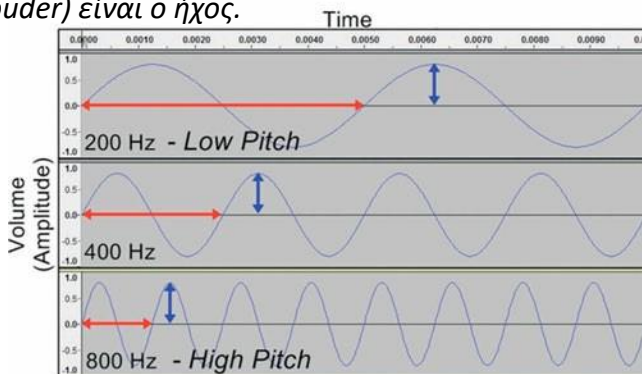
## Κεφάλαιο 5ο - Αναπαράσταση αναλογικού σήματος. Η περίπτωση του ήχου

### 5.1 Ψηφιοποίηση του ήχου με παλμοκωδική διαμόρφωση

Τα ηχητικά κύματα διαδίδονται – χρησιμοποιώντας ως μέσο τον αέρα - με τις μεταβολές της πίεσης (του αέρα) σε σχέση με το χρόνο. Όσο μεγαλύτερη είναι η μεταβολή της πίεσης τόσο αυξάνει η ένταση και όσο πιο γρήγορη είναι η μεταβολή της πίεσης τόσο αυξάνει η συχνότητα και ο ήχος γίνεται πιο οξύς. Αυτές οι αλλαγές του **πλάτους** (ένταση) και της **συχνότητας** (τόνος) είναι οι δύο βασικοί παράγοντες του ήχου. Άλλες συνθήκες όπως ο περιβάλλον χώρος (για παράδειγμα το δωμάτιο) και η πηγή του ήχου, συμβάλουν στη μορφοποίηση του χαρακτήρα του ήχου, επηρεάζοντας το πλάτος και τη συχνότητα που είναι τα δύο βασικά στοιχεία που μας ενδιαφέρουν.



Το πλάτος (*amplitude*) επηρεάζει την ένταση (*volume*) του ήχου. Όσο μεγαλύτερο είναι το πλάτος τόσο πιο δυνατός (*louder*) είναι ο ήχος.



Η συχνότητα (*frequency*) καθορίζει τον τόνο (*pitch*) στον ήχο. Επηρεάζει επίσης το μήκος κύματος (*wavelength*).

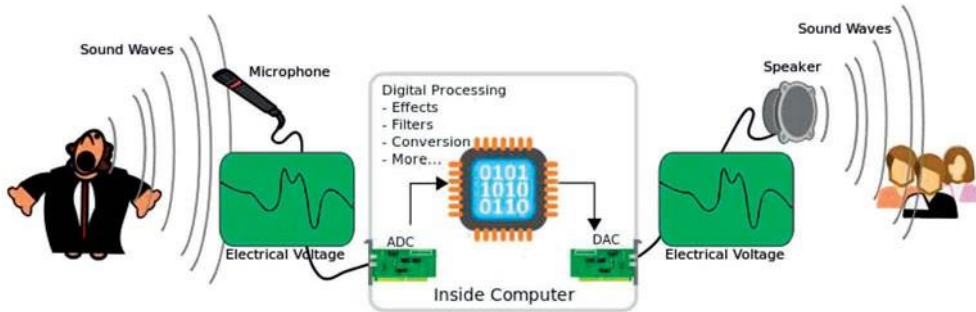
### Αναλογική και ψηφιακή ηχογράφηση

Στην **αναλογική ηχογράφηση**, καταγράφονται οι συνεχείς μεταβολές του αναλογικού ηχητικού σήματος. Για παράδειγμα, οι μεταβολές της πίεσης του αέρα που διεγείρουν τη μεμβράνη ενός μικροφώνου, μετατρέπονται σε μεταβαλλόμενη ηλεκτρική τάση. Οι μεταβολές στην τάση (*electrical voltage variations*) αντιπροσωπεύουν το πλάτος, η ταχύτητα των μεταβολών αντιπροσωπεύει τη συχνότητα και μπορούν να καταγραφούν σε ένα μέσο, όπως η μαγνητική ταινία. Στην αναπαραγωγή, μια συσκευή εξόδου (για παράδειγμα **μεγάφωνο** ή **ακουστικά**) μετατρέπει τη μεταβαλλόμενη ηλεκτρική τάση της πηγής (μαγνητική ταινία) σε μεταβολή της πίεσης του αέρα (ηχητικό κύμα).



**Πως όμως μπορεί να γίνει ψηφιακή ηχογράφηση με το μικρόφωνο;**

Η ψηφιοποίηση του αναλογικού σήματος που μας δίνει το μικρόφωνο, γίνεται με ειδικά ολοκληρωμένα κυκλώματα που λέγονται **αναλογοψηφιακοί μετατροπείς (Analog to Digital Converter, ADC)**. Ο ήχος από αναλογικός, μετατρέπεται σε ψηφιακό, και αφού αποθηκευτεί (digital storage), μπορεί να επεξεργαστεί και να μεταφερθεί ή να αναπαραχθεί. Η αντίστροφη διαδικασία



(αναπαραγωγή) γίνεται με χρήση ενός αντίστοιχου ολοκληρωμένου κυκλώματος που λέγεται **ψηφιοαναλογικός μετατροπέας (Digital to Analog Converter, DAC)**, ώστε το μεγάφωνο να αναπαράγει τον ήχο.

Η διαδικασία που ακολουθεί ένα ADC είναι η ακόλουθη:

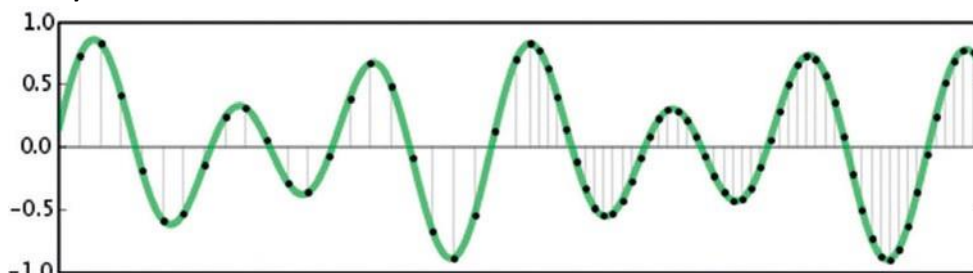
### α) Δειγματοληψία (sampling)

Στο στάδιο αυτό λαμβάνονται δείγματα από το αναλογικό σήμα σε τακτά χρονικά διαστήματα. Από το άπειρο πλήθος τιμών του συνεχούς αναλογικού σήματος, κρατάμε ένα πεπερασμένο σύνολο τιμών σε διακριτές χρονικές στιγμές.

Ο **ρυθμός δειγματοληψίας (Sampling Rate, SR)** είναι ο αριθμός των δειγμάτων ανά δευτερόλεπτο, μετριέται σε KHz και καθορίζει την ποιότητα του σήματος. Σύμφωνα με το **θεώρημα Nyquist-Shannon**, η συχνότητα δειγματοληψίας πρέπει να είναι τουλάχιστον διπλάσια από τη μέγιστη συχνότητα ( $f_{\text{δειγμα}} \geq f_{\text{max}}$ ) που περιέχεται στο σήμα (ώστε ο ψηφιακός ήχος να έχει παρόμοια ποιότητα με τον αναλογικό).

Καθώς στους ανθρώπους, το **ακουστικό φάσμα** (περιοχή ακουστών συχνοτήτων) εκτείνεται κατά προσέγγιση από τα **20 Hz έως τα 20 KHz** (20.000 Hz), το SR πρέπει να είναι τουλάχιστον 40000 Hz. Στο **CD ο ρυθμός δειγματοληψίας είναι 44,1 KHz**, δηλαδή γίνεται μια μέτρηση του πλάτους του σήματος 44.100 φορές το δευτερόλεπτο.

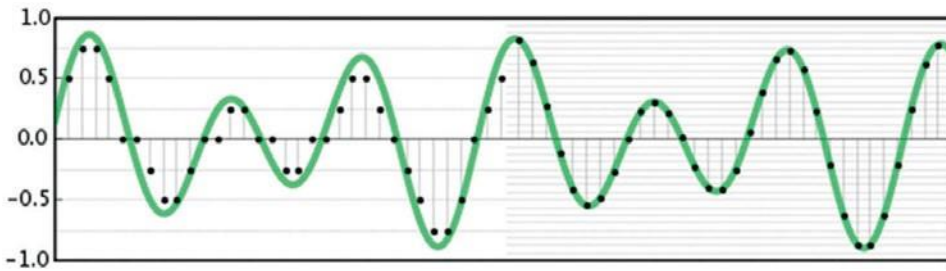
Το σήμα που προκύπτει με αυτό τον τρόπο, αποτελείται από ένα σύνολο παλμών σταθερής χρονικής διάρκειας, οι οποίοι έχουν μεταβλητό πλάτος ανάλογα με τις τιμές του αρχικού σήματος (άπειρες πιθανές τιμές) και είναι γνωστό ως **Διαμόρφωση Πλάτους Παλμών (Pulse Amplitude Modulation, PAM)**.



Σύγκριση χαμηλού ρυθμού δειγματοληψίας με υψηλό ρυθμό δειγματοληψίας

### β) Κβάντιση (Quantization)

Το δεύτερο στάδιο είναι η κβάντιση, δηλαδή η αντιστοίχιση (στρογγυλοποίηση) κάθε δείγματος, στην κοντινότερη τιμή από ένα **πεπερασμένο** πλήθος από **προκαθορισμένες** στάθμες. Με τον τρόπο αυτό, επιβάλλουμε **διακριτές τιμές πλάτους για κάθε δείγμα**, δηλαδή περιορίζουμε τις άπειρες πιθανές τιμές του πλάτους. Μετά το πέρας της κβάντισης το σήμα είναι πλέον ψηφιακό (αλλά όχι απαραίτητα δυαδικό).



### γ) Κωδικοποίηση

Το τελικό στάδιο για την ψηφιοποίηση ενός αναλογικού σήματος είναι η κωδικοποίηση, δηλαδή η αντιστοίχιση κάθε διακριτής στάθμης πλάτους με ένα δυαδικό αριθμό, ώστε να είναι δυνατή η μεταφορά, επεξεργασία και αποθήκευση του σήματος σε υπολογιστικά συστήματα. Η πιο συχνά χρησιμοποιούμενη μέθοδος κωδικοποίησης, είναι η **παλμοκωδική διαμόρφωση (Pulse Code Modulation, PCM)**. Ο αριθμός των δυαδικών ψηφίων (bit) που διατίθενται για την αποθήκευση κάθε στάθμης καλείται **εύρος δείγματος (sample width ή bit depth)**. Το εύρος δείγματος καθορίζει το πλήθος των σταθμών κβάντισης. Για παράδειγμα, ένα ηχητικό σήμα με **εύρος δείγματος 16 bit** (ποιότητα CD) μπορεί να περιέχει 65.536 ( $=2^{16}$ ) στάθμες.

Μπορούμε να έχουμε μια αρχική εκτίμηση του **μεγέθους του αρχείου** (ασυμπίεστου) από τον παρακάτω τύπο:

$$\text{Μέγεθος (byte)} = [\text{Κανάλια (1 ή 2)} \times \text{Ρυθμός δειγματοληψίας (Hz)} \times \text{Εύρος Δείγματος(bit)} \times \text{Διάρκεια (sec)}] / 8$$

Έτσι, για ηχογράφηση ψηφιακού στερεοφωνικού ήχου επιπέδου CD, διάρκειας 5 λεπτών, με ρυθμό δειγματοληψίας 44.1 KHz και εύρος δείγματος 16 bits, απαιτείται αποθηκευτικός χώρος περίπου 52,92 MB.

### 5.2 Τύποι αρχείων ήχου - Συμπίεση

Ο **μορφότυπος (format)** ενός αρχείου περιέχει τη δυαδική πληροφορία και τον τρόπο με τον οποίο κωδικοποιείται. Υπάρχουν πολλοί μορφότυποι αρχείων ήχου, καθένας με τα δικά του πλεονεκτήματα ή μειονεκτήματα, ανάλογα τη σχέση μεταξύ του μέγεθους του αρχείου και της ποιότητας του ήχου.

- Ένα ασυμπίεστο αρχείο ήχου, αναπαράγει τον ήχο με τέλεια ποιότητα, αλλά καταλαμβάνει πολύ αποθηκευτικό χώρο (περίπου 10 MB ανά λεπτό)
- Το αντίστοιχο συμπιεσμένο με μη απωλεστική συμπίεση αρχείο ήχου, καταλαμβάνει περίπου το μισό αποθηκευτικό χώρο, διατηρώντας την ίδια ποιότητα ήχου.
- Όμως οι πιο κοινοί μορφότυποι αφορούν τα συμπιεσμένα με απωλεστική συμπίεση αρχεία ήχου. Ο ήχος είναι γενικά δύσκολο να συμπιεστεί γιατί η ακοή είναι πιο ευαίσθητη στις αλλοιώσεις του ήχου σε σχέση με την όραση. Στην μουσική, οι απαιτήσεις ποιότητας είναι μεγαλύτερες και το εύρος ζώνης της είναι σημαντικά μεγαλύτερο από αυτό της ανθρώπινης φωνής. Τα προγράμματα απωλεστικής συμπίεσης παίρνουν ένα αρχικό ασυμπίεστο αρχείο (π.χ. .wav, .au, .snd, .aiff) και μειώνουν το μέγεθος του πετώντας κάποια πληροφορία που εκείνα θεωρούν άχρηστη ή μη ζωτικής σημασίας για το τελικό αποτέλεσμα. Το συμπιεσμένο αρχείο που παράγουν είναι πολύ μικρότερο σε μέγεθος και χαμηλότερης ποιότητας ήχου σε σχέση με το αρχικό.

### Μορφότυποι ασυμπίεστου αρχείου ήχου (Uncompressed audio file formats)

Ο μορφότυπος **WAVE** (Waveform Audio File Format ή **WAV**) (πρότυπο αποθήκευσης ψηφιακού ήχου για Windows) είναι μορφότυπος αρχείου ήχου (audio container formats) βασισμένος στην κωδικοποίηση PCM, με κάποιες μικρές αλλαγές στην αποθήκευση δεδομένων. Συνήθως χρησιμοποιούνται για την αποθήκευση ασυμπίεστου (PCM) στερεοφωνικού (2 καναλιών) αρχείου ήχου, με δειγματοληψία 44,1 KHz στα 16 bit (ποιότητα CD).

**Μορφώτυποι συμπιεσμένου (με μη απωλεστική συμπίεση) αρχείου ήχου (Lossless compressed audio file formats) : Mpeg m4a**

**Μορφώτυποι συμπιεσμένου (με απωλεστική συμπίεση) αρχείου ήχου (Lossy compressed audio file formats)**

**.MP3 (MPEG Audio Layer 3)** : Συμπιεσμένο αρχείο που αναπτύχθηκε από την Movie Picture Experts Group. Μπορεί να προσφέρει ποιότητα παρόμοια με του CD (stereo 16-bit) σε ένα αρχείο περίπου στο 1/10 του μεγέθους ενός .WAV ή ενός .AIF. Η ποιότητα του εξαρτάται από το **bitrate** που χρησιμοποιήθηκε κατά την συμπίεση του. Τυπικές τιμές είναι 128kbps, 160kbps και 192kbps. Μεγαλύτερες τιμές bitrate έχουν ως αποτέλεσμα μεγαλύτερα αρχεία.

**.AAC (Advanced Audio Coding File)**: Συμπιεσμένο αρχείο ήχου που σε αντίθεση με το MP3, προσφέρει δυνατότητες βελτίωσης της ποιότητας όπως αποδοτικότερη κωδικοποίηση, απλούστερα φίλτρα και καλύτερο χειρισμό των συχνοτήτων άνω των 16KHz.

**.WMA (Windows Media Audio File)**: Αρχείο ήχου συμπιεσμένο με το Windows Media το οποίο κατασκευάστηκε από την Microsoft και μπορεί να μετατραπεί σε άλλες πιο standardized μορφές. Χρησιμοποιείται κυρίως για την αναπαραγωγή μουσικής στο Διαδίκτυο.

#### 5.4 Το πρότυπο MIDI

Η **τεχνολογία MIDI** είναι γύρω μας. Βρίσκεται μέσα σε ηλεκτρονικά μουσικά όργανα, υπολογιστές, ταμπλέτες, έξυπνα τηλέφωνα, μεικτές ήχου, κλπ. Τη **μουσική MIDI** τη συναντούμε σε ταινίες, τηλεόραση, βιντεοπαιχνίδια. Οι **MIDI συνθετητές (Synthesizers)** χρησιμοποιούνται για παραγωγή μουσικής σε συνδυασμό με υπολογιστές που έχουν εγκατεστημένα **προγράμματα μουσικής εγγραφής "Sequencers"** και/ή DAW (Digital Audio workstations).

Το **MIDI (Musical Instrument Digital Interface, Ψηφιακή Διασύνδεση Μουσικών Οργάνων)** είναι το παλαιότερο (1983) και το πλέον διαδεδομένο πρωτόκολλο μουσικής επικοινωνίας, το οποίο επιτρέπει σε ηλεκτρονικά μουσικά όργανα (π.χ. *synthesizers, drum machines*), υπολογιστές και άλλο ηλεκτρονικό εξοπλισμό (π.χ. *MIDI controllers, sound cards, samplers*) να επικοινωνούν και να συγχρονίζονται μεταξύ τους ελέγχοντας το ένα το άλλο.

π.χ. για την ηλεκτρονική χορευτική μουσική (*Electronic Dance Music, EDM*) και τους *DJs*, υπάρχουν ειδικευμένοι *MIDI* ελεγκτές (*MIDI Controllers*) που δημιουργούν "*beats*" και "*loops*" και ελέγχουν το φωτισμό.

Η πληροφορία που μεταφέρεται δεν αποτελεί σήμα ήχου αλλά **μηνύματα** (οδηγίες εκτέλεσης) που αφορούν το ποια νότα πρέπει να παιχθεί, για πόσο χρόνο και ένταση και από ποιο μουσικό όργανο. Επομένως η "**γλώσσα**" **MIDI** ορίζει τον τρόπο αναπαραγωγής της μουσικής με τρόπο παρόμοιο με αυτόν μιας παρτιτούρας. Τα δεδομένα MIDI μπορούν να υποστούν επεξεργασία μέχρι την ελάχιστη λεπτομέρεια (*αλλαγή μιας νότας, αλλαγή του tempo, αλλαγή των μουσικών οργάνων που συμμετέχουν*) με τρόπους που είναι αδύνατον να εφαρμοστούν σε ένα αρχείο ψηφιοποιημένου ήχου.

Τα μηνύματα MIDI μεταφέρονται ηλεκτρονικά ανάμεσα σε συνδεδεμένες συσκευές οι οποίες είναι συμβατές με το πρωτόκολλο MIDI, ή εναλλακτικά αποθηκεύονται σε αρχεία συγκεκριμένου μορφώτυπου (**Standard MIDI files - με επέκταση .mid**) για να αναπαραχθούν από συμβατό λογισμικό ή υλικό.

Επομένως το ηχητικό αποτέλεσμα **εξαρτάται** από τη συσκευή αναπαραγωγής (γεννήτρια ήχου) που θα μετατρέψει τα μηνύματα MIDI σε ήχο (μειονέκτημα της τεχνολογίας MIDI). Ας θυμηθούμε ότι ο ψηφιοποιημένος ήχος είναι ανεξάρτητος της συσκευής που χρησιμοποιείται για την αναπαραγωγή και ακούγεται ίδιος κάθε φορά που αναπαράγεται.

Πλεονεκτήματα του MIDI: υπάρχει μεγαλύτερη ευελιξία στην επεξεργασία της μουσικής MIDI ενώ απαιτείται μικρότερος αποθηκευτικός χώρος (ένα αρχείο MIDI είναι από 200 έως φορές μικρότερο από ένα αρχείο ψηφιοποιημένου ήχου ποιότητας CD).

Μειονεκτήματα: υπάρχει υπολογιστικό κόστος για να μετατραπεί η συμβολική αναπαράσταση MIDI σε ακουστή κυματομορφή σε πραγματικό χρόνο και το αποτέλεσμα εξαρτάται από τη συσκευή εξόδου και συνήθως είναι υποδεέστερο της ψηφιοποιημένης μουσικής.

### Προγράμματα δημιουργίας και επεξεργασίας παρτιτούρας

- **MuseScore**, λογισμικό σύνθεσης και επεξεργασίας μουσικού κειμένου ανοιχτού κώδικα για Windows, MacOS και Linux (<https://musescore.org/el/>, <https://goo.gl/r2Oj7D>)

### 5.5 Σύνθεση και αναγνώριση φωνής (TTS, STT)

Η ανάπτυξη των τεχνολογιών σύνθεσης και αναγνώρισης φωνής είναι τέτοια που επιτρέπει, σε σημαντικό βαθμό, την αυτοματοποίηση της φωνητικής επικοινωνίας. Οι εφαρμογές τεχνολογίας φωνής είναι πλέον αρκετά ώριμες, ώστε να είναι πλέον εύκολη η υλοποίηση αυτόματων συστημάτων πληροφόρησης σε πολύ σύντομο χρονικό διάστημα.

Οι πιο γνωστές υπηρεσίες οι οποίες παρέχονται είναι:

**Φωνο-Αλληλεπιδραστική Απόκριση - Interactive Voice Response (IVR):** Αυτόματα συστήματα που με τη χρήση φωνής αποκρίνονται σε κάποιον που κάλεσε το σύστημα μέσω τηλεφώνου. Ο χρήστης δηλώνει τις επιθυμίες του στο σύστημα φωνητικά ή με τη χρήση των πλήκτρων του τηλεφώνου (Dual-tone multi-frequency signaling, DTMF).

**Αναγνώριση Φωνής - Speech Recognition:** Η αυτόματη διαδικασία αναγνώρισης και επεξεργασίας της ομιλίας.

**Σύνθεση Φωνής - Speech Synthesis:** Συνήθως ο όρος αναφέρεται στη δυνατότητα παραγωγής σήματος φωνής που αντιστοιχεί σε κάποιο κείμενο. Στην περίπτωση αυτή χρησιμοποιείται ο όρος **Σύνθεση Φωνής από Κείμενο (Text-to-Speech)** και αναφέρεται στη διαδικασία μετατροπής κειμένου σε συνθετική φωνή με τη βοήθεια Η/Υ. Ένα σύστημα θα μπορούσε να παράγει συνθετική ομιλία χρησιμοποιώντας:

- Συρραφή (concatenation) προηχογραφημένων λέξεων (περιορισμένο λεξιλόγιο)
- Συρραφή προηχογραφημένων και κατάλληλα αποθηκευμένων φωνημάτων, διφωνημάτων, τριφωνημάτων κλπ., όπου **φώνημα** είναι η μικρότερη ηχητική μονάδα, η οποία διαφοροποιεί τον ένα ήχο από τον άλλο σε οποιαδήποτε ομιλούμενη γλώσσα ή διάλεκτο.
- Χρήση παραμετρικά ελεγχόμενου μοντέλου παραγωγής ομιλίας

### 5.6 Μετάδοση ήχου στο Διαδίκτυο

Το διαδικτυακό ραδιόφωνο είναι η πιο δημοφιλής υπηρεσία για την απευθείας μετάδοση ήχου από το Διαδίκτυο.

Οι πιο δημοφιλείς ραδιοφωνικοί σταθμοί κάνουν αναμετάδοση του σήματος τους και μέσω του Διαδικτύου, παρέχοντας στους ακροατές τους τη δυνατότητα λήψης του σταθμού σε οποιοδήποτε σημείο του πλανήτη, γεγονός που δεν μπορεί να επιτευχθεί μέσα από τη μπάντα των FM. Εκτός από τους παραδοσιακούς ραδιοφωνικούς σταθμούς οι οποίοι βρήκαν το internet ως ένα συμπληρωματικό μέσο μετάδοσης, λειτουργούν πλέον και ραδιοφωνικοί σταθμοί αμιγώς Διαδικτυακοί.

Η διαδικασία μετάδοσης γίνεται από έναν υπολογιστή-πηγή στον οποίο εγκαθίσταται ο **source client**. Ο source client - ο οποίος εγκαθίσταται στο χώρο του ραδιοφωνικού σταθμού και αποτελεί το λογισμικό το οποίο είναι υπεύθυνο να κωδικοποιεί τα ηχητικά δεδομένα στην κατάλληλη μορφή που υποστηρίζει η εκάστοτε τεχνολογία streaming (π.χ. *mp3, Ogg Vorbis, AAC, WMA, Real Audio*). Το λογισμικό αυτό έχοντας ανοιχτή τη σύνδεση με τον **εξυπηρετητή μετάδοσης πολυμεσικών ροών (media streaming server)** αποστέλλει τα κωδικοποιημένα ηχητικά δεδομένα προς αυτόν. Στη συνέχεια ο εξυπηρετητής αναμεταδίδει τα ληφθέντα ηχητικά δεδομένα στους τελικούς χρήστες χρησιμοποιώντας τα πρωτόκολλα TCP ή UDP. Στην ουσία ο ραδιοφωνικός σταθμός μεταδίδει ανά πάσα στιγμή τη ροή στον εξυπηρετητή και εκείνος με τη σειρά του αναμεταδίδει την εισερχόμενη ροή σε πολλαπλούς ταυτόχρονους ακροατές.

Ο χρήστης-ακροατής απαιτείται να έχει εγκατεστημένο ένα **λογισμικό αναπαραγωγής ροής** (είτε ως plugin στο πρόγραμμα πλοήγησης του στο Διαδίκτυο είτε ως αυτόνομο λογισμικό) με το οποίο συνδέεται με τον εξυπηρετητή μετάδοσης ροών (media streaming server) προκειμένου να λάβει και να αναπαράγει τη ροή.