

RAPPORT UIT-R BT.2044

Temps de propagation aller-retour tolérable pour les inserts de programmes radiophoniques et de télédiffusion – Contexte et fondements

(Question UIT-R 35/6)

(2004)

TABLE DES MATIÈRES

	<i>Page</i>
1	Champ d'application..... 2
2	Références bibliographiques et Bibliographie..... 2
3	Termes et définitions..... 3
4	Abréviations et sigles..... 3
5	Origines du retard et de l'écho..... 4
5.1	Contexte..... 4
5.2	Rappel historique..... 4
5.3	Conception et configuration des systèmes et mode d'exploitation..... 5
5.4	Facteurs liés à la conception et la configuration des systèmes qui sont à l'origine du retard ou de l'écho..... 6
5.4.1	Processus de codage/décodage..... 6
5.4.2	Temps de propagation..... 6
5.5	Valeur du retard selon le support de transmission..... 7
5.6	Valeur du retard dans les codecs vidéo..... 7
5.7	Valeur du retard dans les codecs audio..... 8
5.8	Symétrie des circuits..... 10
5.9	Origine de l'écho dans la pratique..... 10
5.10	Retard entre le son et l'image..... 10
6	Effets du retard et de l'écho..... 10
6.1	Retard audio et retard vidéo..... 10
6.2	Variables ayant une incidence sur les perturbations..... 11
6.3	Echos et fusion..... 11
6.4	Types de perturbations et effets de ces perturbations..... 12
6.5	Adaptation..... 12
7	Résumé..... 12

1 Champ d'application

Il s'agit dans le présent Rapport d'étudier les effets du retard et le niveau de l'écho dans une boucle de retour audio dans le contexte de la production de programmes de radiodiffusion. Les effets du retard entre le son et l'image sont aussi examinés. Les effets additionnels de la réverbération et du bruit dans l'environnement d'écoute ne sont pas pris en compte et on suppose qu'il n'y a pas de perte importante de qualité du signal liée au système de contrôle.

2 Références bibliographiques et Bibliographie

- [1] ETSI [juillet 1996] Rapport technique ETR 250 Transmission and Multiplexing (TM). Speech communication quality from mouth to ear for 3.1 kHz handset telephony across networks. Institut européen des normes de télécommunication. <http://pda.etsi.org/pda/queryform.asp>.
- [2] ETSI [janvier 1996] Rapport technique ETR 262 Broadband integrated services digital network (B-ISDN). Asynchronous transfer mode (ATM). Video on demand (VOD) network aspects. Institut européen des normes de télécommunication.
- [3] ETSI [avril 1996] Rapport technique ETR 275 Transmission and Multiplexing (TM). Considerations on transmission delay and transmission delay values for components on connections supporting speech communication over evolving digital networks. Institut européen des normes de télécommunication. <http://pda.etsi.org/pda/queryform.asp>.
- [4] Recommandation UIT-R BT.1359 – Synchronisation relative du son et de l'image en radiodiffusion.
- [5] ETSI [février 1992] ETSI/GSM Recommandation 06.10 Version 3.2.0. GSM Full Rate Speech Transcoding. Institut européen des normes de télécommunication. <http://pda.etsi.org/pda/queryform.asp>.
- [6] DAVIS, P. [1995] A tutorial on MPEG audio compression. *IEEE Multimedia* pp 60-74.
- [7] ONVURAL, R. [1994] *Asynchronous transfer mode networks: performance issues*. Artech, Boston.
- [8] ETSI [février 1999] TR 100 815 V1.1.1 Digital video broadcasting; guidelines for the handling of Asynchronous transfer mode signals in DVB systems. Institut européen des normes de télécommunication.
- [9] C.S0014-0 [décembre 1999] Version 1.0 Enhanced Variable Rate Codec (EVRC) (CDMA2000 specification). Third Generation Partnership Project 2 (3GPP2) EIA/TIA.
- [10] SIU-WAH WONG [1991] An Evaluation of 6.4 kbit/s Speech Codecs for Inmarsat-M System. *IEEE*, p. 629-632.
- [11] Recommandation UIT-R BT.1377 Indication des temps de traitement des appareils audio et vidéo.
- [12] EVEREST, F.A. [1994] *The Master Handbook of Acoustics*. TAB Books/McGraw-Hill.
- [13] Recommandation UIT-T G.114 (2003) – Temps de transmission dans un sens.
- [14] Recommandation UIT-T G.131 (2001) – Dégradations de la transmission dues au traitement vocal.
- [15] ETSI [mars 1998] ES 200 677 V1.2.1 Public Switched Telephone Network (PSTN). Requirements for handset telephony. Institut européen des normes de télécommunication.

- [16] Lochner, J.P.A. et Burger, J.F. [1958] The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech. *Acustica*, Vol. 8, 1 p. 1-10.
- [17] Meyer, E. et Schodder, G.R. [1952] Über den Einfluß von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache. *Nachr. Akad. Wiss.*, Göttingen, 6.
- [18] Church, S. On beer & audio coding – why something called AAC is cooler than a pilsner & how it got to be that way: <http://www.broadcastpapers.com/audio/TelosAAC08.htm>.

3 Termes et définitions

- Latence: Dans les systèmes à commutation par paquets, partie du retard due à l'acheminement des paquets et à leur mise en file d'attente.
- Gigue: Variation de délai pendant la transmission dans un système à commutation par paquets.
- Trame I: Première trame d'un groupe d'images MPEG.
- Trame B: Trame ayant subi une interpolation bidirectionnelle et située entre une trame I et une trame P dans un groupe d'images MPEG.
- Trame P: Trame ayant subi une interpolation dans le sens aller dans un groupe d'images MPEG.
- Effet local: Partie atténuée du signal audio transmis qui revient vers l'émetteur. Il peut être voulu dans les téléphones. Il est causé par un hybride deux-quatre fils asymétrique ou par une terminaison de ligne inadaptée.
- Système à 2 fils: Système dans lequel les signaux sont acheminés dans les deux sens sur la même ligne de transmission.
- Système à 4 fils: Système dans lequel les signaux sont acheminés sur une ligne de transmission distincte, pour chaque sens de transmission.
- Hybride audio: Circuit audio différentiel utilisé pour effectuer la conversion entre un système à 2 fils et un système à 4 fils. Il utilise généralement un transformateur audio à bobinage séparé dans une configuration en pont ou son équivalent en circuits actifs.

4 Abréviations et sigles

- ms: milliseconde (10^{-3} seconde)
- µs: microseconde (10^{-6} seconde)
- ns: nanoseconde (10^{-9} seconde)
- codec: codeur/décodeur
- ETSI: Institut européen des normes de télécommunication (*European telecommunication standards Institute*)
- UIT: Union internationale des télécommunications
- MPEG: Groupe d'experts pour les images animées (*moving picture experts group*)
- fps: Images par seconde (*frames per second*)

5 Origines du retard et de l'écho

5.1 Contexte

Les systèmes de communications en boucle sont couramment utilisés pour la production de programmes de radiodiffusion, lorsque les sites sont nombreux ou éloignés, en particulier pour les programmes en direct. La boucle peut servir à établir la connexion entre l'intervieweur en studio et l'invité qui se trouve sur un site éloigné ou être utilisée comme liaison retour entre le studio et un présentateur se trouvant à un autre endroit lointain. Théoriquement, chaque flux distinct de signaux audio ou vidéo ne devrait circuler que dans un seul sens sur cette boucle. Dans la réalité toutefois, il est fréquent que le signal audio revienne à sa source par phénomène de fuite dans la boucle, en particulier lorsqu'on utilise des téléphones. En raison de la distance et/ou du codage/décodage nécessaire pour comprimer le signal numérisé en vue de sa transmission ou le reconstituer au moment de la réception, le signal peut subir un retard important. Le retard et le phénomène de fuite dans la boucle audio se conjuguent pour produire l'écho.

Etant donné que les contraintes de synchronisation ont toujours été plus strictes pour les images que pour le son, il n'était pas courant d'utiliser de longues boucles pour transmettre les signaux vidéo. Avec l'arrivée de la vidéo numérique, la synchronisation s'est simplifiée, les équipements sont devenus moins volumineux et la transmission de signaux vidéo sur de longues boucles risque de devenir plus fréquente. A l'heure actuelle, il est encore rare que les signaux vidéo empruntent la totalité de la boucle pour revenir à la source. En raison de la synchronisation nécessaire entre l'image et le son, le retard relatif de l'image reste un facteur limitatif dans une boucle de communications audio-vidéo.

Avec des boucles longues, le retard et l'écho peuvent poser de gros problèmes, qu'il s'agisse du maintien de la qualité technique du signal ou de la gêne causée aux participants. Ces deux phénomènes ont des incidences sur les opérations de production.

Le retard et l'écho sont des problèmes distincts mais liés. Le retard est dû à plusieurs facteurs, essentiellement le codage/décodage et la longueur du trajet entre le point d'origine et le point de destination¹. L'écho est dû au phénomène de fuite dans une boucle. Plus le retard est important, plus l'écho qui lui est associé est perceptible et gênant.

5.2 Rappel historique

Le phénomène de gêne provoqué par le retard relatif du son et par l'écho est un phénomène bien connu dans de grands auditoriums et sur les terrains de sport. Le problème n'est pas nouveau. Dans le domaine des communications électroniques, il est apparu pour la première fois avec la téléphonie et une grande partie de nos connaissances sur les origines et les effets du retard en radiodiffusion vient de l'expérience acquise dans ce domaine. Il est donc utile d'avoir une certaine connaissance des origines et des effets du retard en téléphonie ainsi que des stratégies qui ont été mises en oeuvre pour régler le problème. La synchronisation du son et de l'image n'est pas non plus un problème nouveau et l'expérience acquise dans le domaine du cinéma nous a beaucoup aidés.

Avec les premiers systèmes téléphoniques, on avait besoin pour chaque ligne téléphonique d'une paire de fils distincte, pour établir la connexion non seulement avec le central local mais aussi entre les centraux. Le coût des paires de fils était donc une partie importante du coût total de

¹ Le trajet du signal peut être beaucoup plus long que la distance directe ou géodésique entre les deux points. La différence entre la longueur du trajet du signal et la distance géodésique est particulièrement importante lorsque des satellites géostationnaires sont mis à contribution.

l'infrastructure. Grâce à l'utilisation d'un transformateur «hybride» audio (essentiellement un système en pont permettant de combiner ou de dissocier des circuits symétriques), une seule paire de fils (système à deux fils) suffisait pour acheminer les signaux dans les deux sens, ce qui divisait par deux le coût des transmissions sur un système à 4 fils.

Dans ce système, les fonctions de combinaison et de dissociation présentaient des imperfections, car il fallait une bonne adaptation d'impédance avec la ligne côté émission et une ligne avec une impédance de terminaison convenable côté réception. Les caractéristiques électriques des lignes de transmission filaires varient en fonction de la fréquence, de la distance et des conditions atmosphériques (les premiers téléphones comme le télégraphe utilisaient des lignes aériennes) et les premiers hybrides ne pouvaient être adaptés que pour une caractéristique de ligne fixe. Le phénomène de fuite par le transformateur hybride au niveau du combiné, appelé effet local, était donc inévitable en raison à la fois d'une désadaptation d'impédance locale et de l'écho sur la ligne. L'amplitude et la phase de l'effet local varient avec la fréquence et le son qui parvient à l'oreille lorsqu'il y a une légère désadaptation, au niveau de l'hybride, est désagréable et déstabilisant. Si on noie la fuite due à une adaptation non uniforme dans une petite fuite locale large bande, la voix du locuteur telle qu'elle est perçue par l'oreille est plus uniforme et donne l'impression au locuteur que le système fonctionne correctement, du moins à l'extrémité proche. On a donc l'habitude depuis longtemps de tolérer une petite fuite entre les deux demi-boucles audio dans les systèmes à 2 fils.

L'écho provenant de l'extrémité distante d'une ligne caractérisée par une désadaptation d'impédance était un problème relativement mineur dans les premiers systèmes téléphoniques car la longueur des lignes ne dépassait pas 30 km environ. Lorsqu'on a pu amplifier les signaux et que la téléphonie longue distance est devenue possible, le problème de l'écho s'est aggravé, non seulement parce que les points sur la ligne où il pouvait y avoir une désadaptation d'impédance² étaient plus nombreux mais aussi parce que le délai d'écho s'est beaucoup allongé et l'écho est devenu plus perceptible. L'annulation de l'écho est donc devenue une préoccupation majeure pour les compagnies téléphoniques et reste un élément important dans le maintien de la qualité des lignes longue distance.

Avec le multiplexage et la numérisation, les systèmes téléphoniques modernes sont de fait des systèmes 4 fils entre centraux et, dans bien des cas³, cette configuration est conservée jusqu'au combiné. A terme, la numérisation et les circuits 4 fils (ou leur équivalent optique) s'étendront vraisemblablement à tous les combinés. L'écho sur la ligne ne sera alors un problème que pour les téléphones à haut-parleur.

En radiodiffusion, l'utilisation des systèmes de communications à 2 fils est limitée aux lignes d'ordres (intercom) et aux programmes avec intervention téléphonique d'un participant. Pour les programmes en extérieur, on utilise généralement des lignes distinctes pour le canal retour et pour les contributions aux programmes. L'écho est relativement facile à maîtriser dans ces situations⁴ et c'est le retard plus que l'écho qui est le principal problème pour les concepteurs de systèmes.

5.3 Conception et configuration des systèmes et mode d'exploitation

Le retard et l'écho peuvent être dus au système lui-même comme dans le cas des systèmes téléphoniques ou à sa configuration comme dans le cas du paramètre groupe d'images dans un codeur vidéo MPEG ou bien encore à des insuffisances au niveau technique ou opérationnel, par

² Pour une ligne téléphonique analogique longue distance, on a besoin de répéteurs environ tous les 30 km. En chacun de ces points, il peut y avoir une désadaptation d'impédance de la ligne et, en conséquence, un écho.

³ Par exemple, des systèmes PABX numériques connectés numériquement au central local.

⁴ Il n'est pas pour autant garanti que l'écho est maîtrisé.

exemple le fait de ne pas se préoccuper des fuites entre haut-parleurs ou casques de retour et micros ouverts. Avec un système bien conçu et bien configuré et une bonne méthode d'exploitation, on pourra contenir dans des limites raisonnables le retard et l'écho sur les boucles.

5.4 Facteurs liés à la conception et à la configuration des systèmes qui sont à l'origine du retard ou de l'écho

Le retard dans les systèmes audio/vidéo est dû essentiellement au processus de codage/décodage et à la propagation du signal.

5.4.1 Processus de codage/décodage

Dans les systèmes numériques, le processus comprend deux étapes, à savoir la conversion analogique-numérique du signal puis la compression des données. Pour le décodage c'est l'inverse: le signal subit une décompression puis une conversion numérique-analogique.

La conversion analogique-numérique ou l'inverse sont des processus relativement rapides qui se font généralement dans l'intervalle entre échantillons. Pour les programmes radiophoniques, cet intervalle est d'environ 1/48 kHz ou 21 μ s. Pour la téléphonie audio, il est d'environ 1/8 kHz ou 125 μ s et pour la vidéo à définition normale d'environ 1/13 MHz ou 77 ns.

La compression des données est utilisée pour réduire la largeur de bande nécessaire et elle est presque universellement utilisée en radiodiffusion pour les transmissions sur de longues distances (entre des bâtiments, des localités, des villes ou des pays). En général, le son et l'image dans la chaîne de radiodiffusion sont transmis sous forme numérique compressée pendant la production, l'échange, la distribution et la diffusion de programmes. La compression des données est utilisée pratiquement dans le monde entier pour l'enregistrement même si certaines normes de compression pour l'enregistrement diffèrent de celles pour la transmission.

Les retards dus à la compression des données sont légèrement plus longs que ceux liés à la conversion analogique-numérique ou numérique-analogique. Pour avoir des facteurs de compression élevés, on réduit la redondance entre les trames vidéo en comparant les trames à une trame de référence et en codant uniquement les différences. Le codage vidéo se fait donc sur la base d'un groupe d'images et non trame par trame. Pour la comparaison des trames au stade du codage, on utilise des blocs de 15 trames au maximum avec des sous-groupes de quatre trames au maximum. Pour un groupe d'images de 15 images, le retard minimal est de quatre trames. Le retard de codage total, temps de traitement compris, peut être beaucoup plus important que le nombre minimal de trames tampon et il est habituellement d'au moins sept ou huit trames pour cette configuration. Les retards sont analogues côté décodage.

Le retard dû à la compression des données du son numérisé est du même ordre, même si les techniques de compression sont légèrement différentes de celles utilisées pour la vidéo.

5.4.2 Temps de propagation

Le temps de propagation comprend la composante distance et la composante commutation.

Les temps de propagation liés à la distance sont de l'ordre de 5 μ s/km de trajet dans un guide d'ondes, un câble ou une fibre optique et de l'ordre de 3,3 μ s/km pour les signaux rayonnés. Dans les deux cas, la longueur du trajet peut être analogue à la distance directe point à point (géodésique) ou beaucoup plus importante, si les transmissions se font par satellite. Pour de longues distances, ce retard peut être comparable à celui lié au codage et au décodage. Pour un signal acheminé par voie hertzienne de Terre jusqu'à l'autre bout du globe, le retard lié à la distance est d'environ 100 ms. Par satellite, ce retard, pour la même destination, est d'environ 280 ms dans l'hypothèse où n'intervient qu'un seul satellite géostationnaire (altitude verticale de 36 000 km et de 42 000 km par rapport à l'horizon).

Les temps de commutation ne sont pas généralement perceptibles dans les systèmes à porteuses analogiques ou les systèmes numériques synchrones. Dans les systèmes numériques à commutation par paquets, comme les systèmes fonctionnant en mode de transmission asynchrone (ATM), il y a un temps d'attente à chaque noeud de commutation étant donné que chaque paquet est retenu et que son en-tête est lu avant retransmission dans l'espace suivant disponible sur une ligne adéquate. Il y a donc un retard lié au processus de lecture/d'attribution du trajet et à la mise en file d'attente. Ce retard est appelé latence. Etant donné que le trajet de transmission est attribué de façon dynamique pour chaque paquet dans les systèmes à commutation par paquets, il y a une certaine variabilité du temps de transmission appelée gigue. Les retards associés à la commutation des paquets sont toutefois faibles (quelques ms) par rapport au retard lié à la distance dans le cas de communications longue distance. Ils sont aussi assez uniformes pour un trajet constant: la plus grande partie de la gigue est liée généralement aux variations de la longueur du trajet.

5.5 Valeur du retard selon le support de transmission

Le retard est dû essentiellement à deux raisons: le temps de propagation et le codage/décodage. Le temps de propagation varie en fonction du mode de transmission, par voie hertzienne de Terre ou par satellite, en continu ou par paquets.

- Temps de propagation
 - Trajet satellite: pour des satellites géostationnaires à une altitude d'environ 36 000 km, comme les satellites de radiodiffusion, le temps de propagation sur le trajet, fonction de la distance, se situe entre environ 239 ms lorsque le satellite est exactement au-dessus de la source et du récepteur et environ 281 ms lorsque le satellite est à l'horizon pour la source et le récepteur.
 - pour des satellites en orbite basse comme ceux utilisés dans le système téléphonique mobile Iridium, l'altitude est voisine de 780 km et le retard est proportionnellement moins important. Etant donné que ces satellites ne sont pas géostationnaires, le retard varie en fonction de leur position. Avec une zone de couverture de 1 800 km de rayon pour chaque satellite (il y en a 66 dans le système Iridium), la distance maximale par rapport au satellite est de 2 000 km, ce qui donne un temps de propagation dans un sens d'environ 6,6 ms. Des retards supplémentaires imputables aux liaisons inter-satellites et au traitement à bord du satellite feront augmenter le retard total dans le système Iridium.
 - Systèmes analogiques et RNIS: le temps de propagation pour les systèmes analogiques de Terre et les systèmes RNIS de Terre est d'environ $5 \mu\text{s}/\text{km}$, sur la base d'une vitesse de propagation dans un guide d'ondes optiques ou en cuivre égale à environ $2/3$ de la vitesse de la lumière. On obtient un temps de propagation de 100 ms lorsque le signal est acheminé jusqu'à l'autre bout de la Terre.
 - Systèmes ATM: pour ces systèmes, aux $5 \mu\text{s}/\text{km}$ s'ajoute un temps de commutation (latence) lié à la technique d'acheminement des données par paquets. Le temps de latence maximal est de 150 μs à chaque noeud et il peut y avoir jusqu'à 40 noeuds sur un long trajet de transmission. Dans la pratique, le temps de latence est habituellement de 20 à 40 μs par noeud.

5.6 Valeur du retard dans les codecs vidéo

Pour le codage vidéo en radiodiffusion, on utilise presque partout la compression MPEG pour la transmission longue distance. Le temps de codage en vidéo MPEG dépend du paramètre groupe d'images (GoP) qui dans la plupart des codecs est fixé entre 1 et 15 trames. Il ne dépend pas de la longueur totale du groupe d'images mais de la longueur du sous-groupe contenant la première trame

(trame I) et la première trame P. Le paramètre GoP est fixé à 15 pour maximiser la compression des données. La longueur du premier sous-groupe est alors de 4 trames, avec un temps de codage minimal de 133 ms à 30 fps ou de 160 ms à 25 fps; le temps de codage type est environ le double de ces chiffres. Si la valeur donnée au paramètre GoP est plus faible, le temps de codage minimal peut être ramené, respectivement à 33 ms et 40 ms au détriment de la compression des données (débit binaire élevé) ou la qualité de l'image à un débit binaire constant.

5.7 Valeur du retard dans les codecs audio

Les caractéristiques des codecs audio varient beaucoup selon l'application. En radiodiffusion, on utilise souvent des téléphones fixes ou mobiles pour les contributions aux programmes, en particulier pour les actualités et les programmes avec intervention d'un participant par téléphone. La diffusion audio en continu sur Internet est de plus en plus fréquente dans les services de radiodiffusion mais n'est pas encore beaucoup utilisée pour les contributions aux programmes. Les programmes audio de qualité peuvent aussi être codés en vue de leur transmission avec un codec de qualité radiodiffusion.

Les valeurs types du retard de transcodage pour une série de systèmes sont les suivantes:

- Téléphone mobile GSM [5]: valeur recommandée pour le retard de transcodage: 20 ms au minimum et 30 ms au maximum.
- Téléphone mobile AMRC [10]: valeur recommandée pour le retard de codage: 20 ms au maximum; valeur recommandée pour le retard de décodage: 3 ms au maximum.
- Téléphones par satellite: plusieurs systèmes sont en concurrence, notamment ICO, Iridium, AceS, AMSC-TMI et Inmarsat. Ils peuvent être classés en plusieurs catégories: systèmes en orbite terrestre basse (LEO) (ICO, Globalstar, Iridium et Teledesic), ou systèmes à satellites géostationnaires en orbite autour de la Terre (GEO) (Inmarsat, Satphone, ASC, Thuraya, APMT, EAST). Un système OSG type, comme le système Inmarsat Mini-M, utilise un codage avec excitation multibande évoluée (AMBE) à 4,8 kbit/s, et des retards de codage/décodage comparables à ceux des systèmes de téléphonie mobile de Terre GSM et AMRC. Dans le cas d'un système OSG toutefois, le retard lié à la distance rend le retard de codage/décodage négligeable.
- Codec MPEG de studio.

Les signaux audio pour la radiodiffusion peuvent être intégrés aux signaux vidéo en utilisant un codage MPEG-2 ou être acheminés sur des lignes distinctes RNIS/ATM. Dans le premier cas, ils seront synchronisés avec les signaux vidéo. Dans le second cas, les signaux audio pourront être décompressés et le retard de codage/décodage sera alors négligeable, ou compressés. Les algorithmes de compression varient d'un fabricant à l'autre de sorte que l'on indique les valeurs minimales et les valeurs types pour le retard.

Compte tenu de la nature des signaux audio, un débit binaire minimal pour une qualité du son donnée varierait d'un facteur de 10 ou plus avec le temps. En règle générale, les intervalles de temps où le débit binaire local est très sollicité sont très courts et entourés d'intervalles de temps où le débit binaire est très faible. Dans la plupart des schémas de codage audio modernes, on établit une moyenne du débit binaire sur le temps pour obtenir un débit binaire constant. Selon que le tampon nécessaire pour établir cette moyenne est situé côté codeur ou côté décodeur, le retard de codage ou de décodage est plus long.

Il y a plusieurs options pour la compression des signaux audio MPEG-1: Couche I, Couche II et Couche III. La Couche III est également appelée MP3. Le codage MPEG-2 en plus des Couches I, II et III offre un codage audio évolué (AAC, advanced audio coding). Dans le codage MPEG-4, le codage AAC a été choisi comme codage de base pour les signaux audio naturels et plusieurs nouveaux outils et fonctions lui ont été ajoutés. Pour ce qui est du temps de propagation aller-retour, la version retard faible (LD) du codage AAC est l'extension la plus visible.

Valeurs minimales du retard pour le codage MPEG-1 Couches I, II et III de [6], [18]

Couche	Débit binaire cible (kbit/s)	Taux de compression	Retard minimal théorique (ms)
AAC-LD	64	12/1	20
Couche I	192	4/1	19
Couche II	128	6/1	35
Couche II	64	12/1	59

Codage	Débit binaire (kbit/s)	Taux d'échantillonnage (kHz)	Retard type (ms)
AAC-LD stéréo	128	48	60
AAC-LD mono	64	48	50
AAC stéréo	128	48	172
Couche III stéréo	128	48	326
Couche II stéréo	128	48	224
Couche II stéréo	128	24 (half mode)	398
G.722	64	48	10

Retard de bout en bout (y compris canal RNIS) pour un codec audio disponible sur le marché [18].

Codec	Débit binaire (kbit/s)	Retard algorithmique avec réservoir de bits mis à zéro (ms)	Pleine charge, transmission par salves (ms)	Pleine charge, transmission continue (ms)	Charge à 30%, transmission par salves (ms)	Charge à 30%, transmission continue (ms)
Couche 2	192	34	NA	NA	NA	NA
Couche 3	128	54	118	142	107	131
MPEG-4 AAC	96	55	82	211	63	192
MPEG-4 HE AAC	56	129	184	361	145	322
MPEG-4 AAC-LD	128	20	33	44	24	35

Retard de bout en bout pour les schémas de codage audio actuels.

Les signaux audio MPEG codés avec les signaux vidéo MPEG auront en général le même retard que les signaux vidéo.

5.8 Symétrie des circuits

Il ne faut pas supposer automatiquement que les deux demi-boucles d'une boucle de contribution aux programmes seront identiques pour ce qui est du temps de traitement, du temps de propagation sur le trajet, du retard global ou des phénomènes de fuite. On a l'habitude en radiodiffusion d'utiliser un circuit de haute qualité uniquement sur la ligne affectée aux programmes en extérieur et un circuit de qualité inférieure sur la ligne de retour. Il est également courant d'utiliser un circuit retour uniquement audio même sur des boucles de contribution vidéo. Ces deux éléments auront une incidence sur le retard global et sur la gêne liée au retard et à l'écho telle qu'elle est perçue.

5.9 Origine de l'écho dans la pratique

Un certain retard est inévitable dans un circuit audio mais l'écho peut en général être évité. L'écho peut être dû à la conception ou à une mauvaise utilisation du système. Une bonne utilisation du système ne garantira pas nécessairement qu'il n'y aura pas d'écho sur la boucle, mais l'écho sera réduit au minimum.

L'efficacité d'une bonne utilisation du système dans la limitation de l'écho dépend dans une large mesure de l'écho inhérent à une boucle. Dans un circuit à 2 fils, l'écho intrinsèque est déterminé par la conception du système. Dans un circuit à 4 fils, l'écho imputable au phénomène de fuite du signal audio entre les circuits de retour et les circuits affectés aux programmes peut être limité en:

- utilisant une sortie directe ou partielle («mix-minus» ou «clean feed»); et/ou
- coupant le circuit de retour lorsque les microphones sont allumés, soit par un système de verrouillage si on utilise un interrupteur manuel soit à l'aide d'un circuit activé par la voix («ducker»); et/ou
- contrôlant les casques d'écoute fermés ou les oreillettes afin de limiter les fuites dans les microphones ouverts.

5.10 Retard entre le son et l'image

Le retard entre le son et l'image a été bien étudié et des normes [4] sont en vigueur depuis un certain temps. Le principal problème au niveau de la perception, avec ce type de retard, est la perte de synchronisation labiale. La norme de l'UIT en vigueur spécifie un retard acceptable entre le son et l'image de +25 ms (le son est en avance sur l'image) à -100 ms (l'image est en avance sur le son), retard mesuré au niveau du dispositif de sélection de source finale de programme (en règle générale la régie).

Il est théoriquement facile de resynchroniser le son et l'image en faisant varier le retard, même si l'équipement nécessaire peut être onéreux. Si le son est en avance sur l'image, on peut utiliser une unité de retard audio simple et peu coûteuse pour corriger la synchronisation. Si l'image est en avance sur le son, l'image doit être retardée: l'équipement nécessaire est en général plus coûteux même si ce dispositif peut être disponible dans un serveur vidéo existant.

6 Effets du retard et de l'écho

6.1 Retard audio et retard vidéo

Les boucles de communications vidéo sont utilisées depuis quelques années déjà en visioconférence et en radiodiffusion mais elles représentent une technologie relativement nouvelle comparées aux boucles de communications audio. C'est pourquoi, l'effet subjectif d'un retard important dans une boucle de communications vidéo (de plus d'une trame) n'est pas aussi bien étudié que dans le cas d'une boucle audio. Les informations concernant les effets subjectifs du retard dans les boucles vidéo sont donc relativement peu nombreuses comparées à celles dont on dispose pour les boucles

audio. La présente contribution est donc basée essentiellement sur les informations dont on dispose concernant l'effet subjectif du retard et de l'écho dans une boucle audio. En attendant d'avoir des informations plus détaillées sur les effets subjectifs du retard dans les boucles vidéo, on part de l'hypothèse que le retard audio et l'écho constituent la principale cause des perturbations, qu'ils constituent ou non le facteur limitatif dans le retard global sur une boucle audio-vidéo.

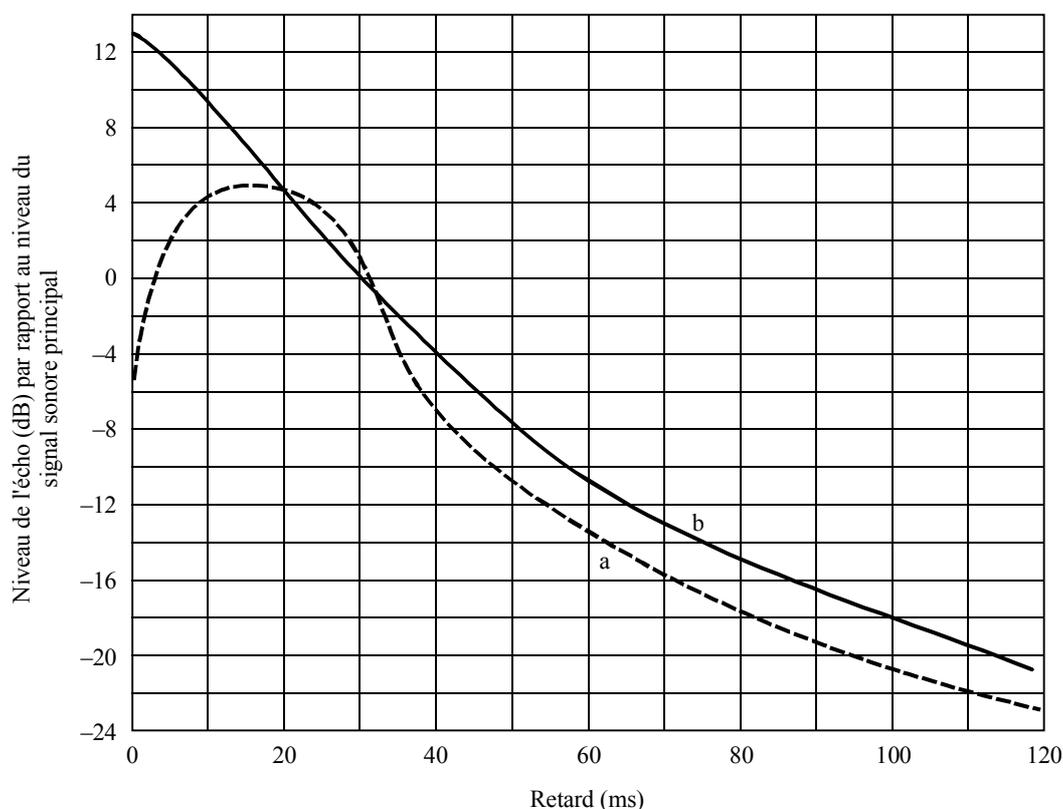
6.2 Variables ayant une incidence sur les perturbations

Les retards dans une boucle audio perturbent les systèmes de communications vocales. Les effets perturbateurs dépendent essentiellement de deux facteurs: la longueur du retard et l'affaiblissement sur le canal retour dans la boucle [1, 3].

6.3 Echos et fusion

Pour des retards inférieurs à la valeur seuil (voir la Fig. 1), l'oreille humaine ne peut pas faire la distinction entre les deux signaux entrants pour la plupart des types d'informations vocales ou musicales. Les deux sons semblent donc «fusionner» en un son plus fort, ce qui a un effet positif à la fois sur l'intensité sonore perçue et sur l'intelligibilité de la parole. L'effet de fusion diminue progressivement avec le retard, en particulier lorsque le retard est supérieur à 30 ms. Lorsqu'il n'y a pas fusion, l'écho devient gênant.

FIGURE 1
Audibilité du signal retardé en présence du signal direct, en fonction du retard et du niveau relatif du signal [16]



Comparaison des seuils des courbes de perception.

Courbes a: Meyer et Schodder (niveau du signal sonore principal 55 Phons)
b: Lochner et Burger (niveau du signal sonore principal 50 Phons)

6.4 Types de perturbations et effets de ces perturbations

Les perturbations imputables au retard et à l'écho sont essentiellement de trois types:

- Retard important sans écho retour significatif. Ce type de retard relève de la catégorie temps de propagation de demi-boucle. Il peut causer des difficultés dans une conversation normale: involontairement les paroles du locuteur et de l'auditeur se chevauchent et les phrases sont interrompues dès que l'on s'aperçoit de ce phénomène. La conversation est ainsi prolongée et hachée, ce qui distrait l'attention. Les retards sur les circuits par satellite (de l'ordre de 240 à 280 ms pour chaque liaison par satellite) sont des causes classiques de ce type de perturbation.
- Echos revenant vers le locuteur. Ce type de retard correspond à un multiple entier (1, 2, ... n) d'un temps de propagation de boucle entière. Ces échos perturbent le mécanisme normal de retour de la parole (cerveau-bouche puis oreille-cerveau), causant des bégaiements et des hésitations. La durée du retard et le niveau de l'écho ont tous les deux des conséquences importantes sur les perturbations; en effet, elles augmentent avec le retard à un niveau constant et diminuent avec le niveau pour un retard constant. Pour des temps de propagation importants (de plus de 240 ms) des affaiblissements de retour aussi faibles que -50 dB peuvent causer des perturbations.
- Echos revenant à l'auditeur. Il s'agit de temps de propagation de 1 boucle $1/2$, 2 boucles $1/2$. Ils perturbent l'auditeur car ils réduisent l'intelligibilité de la parole. Comme avec les temps de propagation de boucle entière, les perturbations s'aggravent avec la longueur du retard. Le problème est courant lors de discours prononcés en public. On parle de «slap echo» pour des réflexions simples et de «flutter echo» (écho tunnel) pour des réflexions multiples. L'audibilité de ce type de retard est illustrée à la Fig. 1 [16]. Pour être inaudibles en tant que sons, les échos devraient être situés au-dessous de cette courbe.

Les effets du délai d'écho et du niveau de l'écho d'une part sur l'intelligibilité de la parole et d'autre part sur la gêne et la difficulté à parler ont été beaucoup étudiés par les autorités des télécommunications [1, 3, 14] et sont relativement bien compris. Les premiers sont probablement mieux compris toutefois que les seconds.

6.5 Adaptation

On appelle adaptation une tolérance apprise ou acquise aux effets d'un stimulus ou d'un phénomène irritant. Il est possible de s'adapter, dans une certaine mesure, à des signaux affectés par l'écho mais l'adaptation possible est en général assez limitée. Il faut aussi parfois un certain temps pour s'adapter. L'adaptation qui peut être apprise et le temps nécessaire pour l'apprendre varieront vraisemblablement beaucoup d'un individu à l'autre.

Etant donné qu'une grande partie du contenu des programmes radiodiffusés, en particulier pour les shows en direct, est fournie par des invités et non par les présentateurs, il convient d'élaborer des lignes directrices en partant de l'hypothèse qu'il n'y aura pas d'adaptation dans le temps disponible.

7 Résumé

Les systèmes de communications en boucle sont couramment utilisés pour la production de programmes de radiodiffusion lorsque les sites sont nombreux ou éloignés, en particulier pour les programmes en direct. La boucle peut servir à établir la connexion entre l'intervieweur en studio et l'invité qui se trouve sur un site éloigné ou être utilisé comme liaison retour entre le studio et un présentateur se trouvant à un autre endroit, lointain. Théoriquement, chaque flux distinct de signaux audio ou vidéo ne devrait circuler que dans un seul sens sur cette boucle. Dans la réalité toutefois, il

est fréquent que le signal audio revienne à sa source par phénomène de fuite dans la boucle, en particulier lorsqu'on utilise des téléphones. En raison de la distance et/ou du codage/décodage nécessaire pour comprimer le signal numérisé en vue de sa transmission ou le reconstituer au moment de la réception, le signal peut subir un retard important.

Le retard et le phénomène de fuite se conjuguent pour produire l'écho. Le retard et l'écho peuvent causer des perturbations qui ont des conséquences négatives pour les participants au programme et pour les téléspectateurs ou les auditeurs. Le retard et l'écho sont à l'origine essentiellement de trois types de perturbations:

- Retard important sans écho retour significatif.
- Echos revenant vers le locuteur.
- Echos revenant vers l'auditeur.

L'effet de l'écho sur l'intelligibilité de la parole est relativement bien compris. L'effet du retard sur le locuteur et sur l'auditeur est moins bien compris. Les lignes directrices proposées ici concernent donc l'audibilité et les effets sur l'intelligibilité de la parole.

Le phénomène de gêne provoqué par le retard relatif du son et par l'écho est un phénomène bien connu dans de grands auditoriums ou sur les terrains de sport. Le problème n'est pas nouveau. Dans le domaine des communications électroniques, il est apparu pour la première fois avec les systèmes téléphoniques à 2 fils et une grande partie de nos connaissances sur les origines et les effets du retard en radiodiffusion vient de l'expérience acquise dans ce domaine. Il est difficile de maîtriser les phénomènes de fuite entre les deux demi-boucles d'une boucle audio dans les systèmes à deux fils et on a donc l'habitude depuis longtemps de tolérer une fuite maîtrisée dans ces systèmes pour que la fuite incontrôlable soit moins perceptible.

Il y a essentiellement deux causes de retard: le temps de propagation et le temps de codage/décodage. Le temps de propagation comprend la composante distance et la composante commutation.

Les temps de propagation liés à la distance sont de l'ordre de 5 $\mu\text{s}/\text{km}$ de trajet dans un guide d'ondes, un câble ou une fibre optique et de l'ordre de 3,3 $\mu\text{s}/\text{km}$ pour les signaux rayonnés. Pour un signal acheminé par voie hertzienne de Terre jusqu'à l'autre bout du globe, le retard lié à la distance est d'environ 100 ms. Par satellite, pour la même destination, il est d'environ 280 ms, dans l'hypothèse où n'intervient qu'un seul satellite géostationnaire.

Compte tenu de la nécessité de la synchronisation entre le son et l'image, le retard relatif de l'image est un facteur limitatif dans une boucle de communications audio-vidéo. Le codage vidéo se fait donc sur la base d'un groupe d'images et non sur des trames individuelles. Pour un groupe d'images (GoP) de 15 images (la plus grande taille couramment utilisée), on obtient la compression des données la plus élevée. Le retard minimal est de quatre trames. Le temps total de codage, y compris le temps de traitement, peut être beaucoup plus important que le nombre minimal de trames tampon et il est habituellement d'au moins 7 ou 8 trames (230-320 ms) pour cette configuration.

Les retards de transcoding audio dans la plupart des systèmes de téléphonie mobile, GSM, AMRC ou téléphone par satellite, sont faibles (3-30 ms). Pour un codage audio de radiodiffusion de qualité supérieure comme le codage MPEG AAC, le retard peut être du même ordre que celui pour le codage vidéo (170-400 ms). Pour des systèmes plus récents comme les systèmes MPEG AAC-LD, ce retard est ramené à environ 60 ms.

Les systèmes ATM ont un temps de commutation (latence) lié à la technique d'acheminement des données par paquets. Le temps d'attente maximal à chaque noeud est de 150 μs et il peut y avoir jusqu'à 40 noeuds sur un long trajet de transmission. Dans la pratique, le temps de latence est habituellement plus proche de 20 à 40 μs par noeud, ce qui se traduit par un temps de propagation relativement faible d'environ 1,6 ms.

L'écho peut être réduit au strict minimum dans un système bien conçu et bien utilisé.

Trois principes généraux devraient être respectés lors de la conception et de la configuration des installations nécessaires pour les boucles de contribution aux programmes:

- limiter le nombre d'étapes de codage/décodage;
- maintenir le retard relatif entre le son et l'image dans une fourchette comprise entre +25 ms et -100 ms;
- limiter le phénomène de fuite sur le circuit retour et le maintenir au-dessous du seuil d'audibilité.

Pour limiter l'écho on peut notamment:

- utiliser une sortie directe ou partielle «mix-minus» ou «clean feed»; et/ou
 - couper le circuit de retour lorsque les micros sont allumés, soit par système de verrouillage si on utilise un interrupteur manuel soit à l'aide d'un circuit activé par la voix («ducker»); et/ou
 - contrôler les casques d'écoute fermés et les oreillettes afin de limiter les fuites dans les microphones ouverts.
-