ETRE DEDANS, ETRE DEVANT
Son & parole à la télévision d'un point de vue cognitiviste

Le téléspectateur oscille entre un positionnement extérieur (*être devant*) et un positionnement intérieur (*être dedans*), avec tous les degrés intermédiaires possibles. Être devant constitue le *modèle du moniteur* (*ça* (me) parle) ; il s'agit de « regarder la TV » avec une attention intermittente (peu importe ce qui passe), ou bien de regarder l'*objet* TV, quand il représente un investissement financier important et exemplifie un exploit technique (*home cinema*). Être dedans constitue le *modèle de la fenêtre* (*il* (me) parle) ; il s'agit cette fois de suivre une émission particulière comme on regarde un film au cinéma, avec le double jeu des identifications, primaires et secondaires (absorption diégétique + confrontation exhibitionniste : on retrouve, comme "formes de sollicitation", pour reprendre les termes du programme du séminaire *Télé-parole*, les deux registres du cinéma primitif tels qu'ils ont été décrits par Tom Gunning). Ou encore, seconde forme de positionnement intérieur interdite cette fois au cinéma, de suivre une émission particulière dans une perspective interactive (téléphoner au standard, télé-achat...).

Ces deux modèles déterminent pour beaucoup l'attitude d'écoute du téléspectateur. Le texte qui suit analyse cette attitude d'écoute à l'aide d'outils inspirés de la psychologie cognitive ; il va de soi qu'une telle démarche est plus efficace en ce qui concerne le modèle de la fenêtre (qui suppose un investissement émotionnel et affectif à base de *cadres*) qu'en ce qui concerne le modèle du moniteur (où des outils en provenance de la sociologie ou de la psychologie sociale auraient davantage de puissance heuristique). On ne prétend donc pas ici modéliser la totalité des comportements téléspectatoriels possibles – entreprise que les contempteurs des approches « scientistes » de la réception ne se lassent pas, bien que tout le monde s'accorde à le penser, de qualifier de chimère -, mais seulement donner quelques pistes et quelques bornes.

SON & PAROLE

L'une des questions posées dans le programme liminaire de ce séminaire est la différence entre son et parole. Du point de vue physique, il n'y en a guère ; on se trouve en présence d'*étiquettes* qui sont collées - en contexte exclusivement, comme en ce qui concerne la musique - sur des signaux acoustiques. Hors contexte, examinés au spectrographe par exemple, ces signaux acoustiques se ressemblent les uns les autres. Verbal *vs.* non-verbal, musical *vs.* non musical, sont des axes d'opposition qui servent à l'homme pour extraire des informations utiles dans le continuum sonore, mais qui fluctuent selon les époques historiques et les lieux géographiques. Ces axes déterminent un travail de filtrage du système perceptif (ce qui n'a pas de valeur distinctive est écarté), et un travail d'interprétation du système cognitif (en fonction de scripts de lecture appris, on peut se livrer à des restaurations phonémiques, ou des prédictions : par exemple on reconnaît mieux un mot s'il est précédé de mots sémantiquement associés). Résultat, la différence est grande entre le signal émis et le signal reçu. Ainsi, la musique tonale et la parole, du strict point de vue du signal physique, sont des *systèmes faux*. Le "clavier bien tempéré" ne recourt pas à la mathématique. Quant au traitement de l'information verbale, il se fait par empans d'une syllabe à la fois ; or la syllabe est la plus petite unité de véhiculage du contraste entre consonnes et voyelles, et sa longueur correspond à la durée de stockage du système sensoriel. Mais nous percevons de façon *discrète*, sous forme de formants (voyelles) et de transitions (consonnes), alors que le signal acoustique est *continu*, et qui plus est contient des chevauchements de sons que littéralement

on n'entend pas ; par ailleurs il n'y a pas d'invariants acoustiques correspondant aux phonèmes ; un même signal acoustique peut être lu de façon différente selon le contexte…

Détail intéressant, l'apprentissage de la parole - et du monde en général d'ailleurs - se fait sur la base de stimuli audiovisuels *synchrones* - le visuel a son mot à dire. Par exemple si une TV diffuse (hors contexte) la syllabe /ba/ mais qu'il voit en même temps une bouche prononcer /ga/, le sujet dira avoir entendu /da/… En fait nous sommes programmés pour n'entendre (au sens de "comprendre") que ce qu'on peut envisager de produire avec notre bouche - c'est ce qui s'appelle la *théorie motrice*. Avant un an l'enfant discrimine tous les phonèmes possibles - même ceux qui ne sont pas pertinents dans sa propre langue, ensuite il se concentre sur sa langue. Par exemple un bébé japonais pourrait distinguer entre /roi/ et /loi/, alors qu'un adulte ne le pourrait plus, l'opposition l/r n'étant pas pertinente en japonais… La seule exception à cette règle de l'apprentissage audiovisuel, que monte par ailleurs en épingle tout le courant des *gender studies* trempé à la psychanalyse selon M. Klein et J. Lacan, serait la *voix de la mère* qui parle ou chante au moment de l'endormissement (invisible raconteuse - *invisible storyteller* – selon le mot de S. Kozloff). Dans cette perspective *gender & psychoanalytic*, une voix off aura d'autant plus d'impact sur le téléspectateur que la seule voix qu'il aura laissée off au cours de son apprentissage du monde - il n'aura pas tourné la tête en direction du visuel correspondant, parce qu'il avait confiance et qu'il était fatigué – aura été celle de la mère. Il est impossible de vérifier expérimentalement la validité de ces modèles psychanalytiques ; par ailleurs la connexion audiovisuelle n'est pas la seule forme d'intermodalité – la voix et le toucher entretiennent ainsi des liens étroits. « Le grain, écho lointain de l'oreille tactile, se mue en caresse ; un simple câble issu du cortex auditif débusque les aspérités des sons et les projette au cortex tactile pour y éveiller la sensation, grenue ou lisse, d'un toucher. Il est parfois si intense qu'il se projette à son tour à la surface de la peau, ondulant en frissons » (Bailblé 1999 : 364).

L'ATTITUDE D'ECOUTE

L'évolution n'ayant pas (encore) équipé son appareil perceptivo-cognitif en fonction des machines de simulation audiovisuelle, le téléspectateur écoute la TV comme il écoute la monde, c'est-à-dire qu'il combine face à une voix les trois types d'écoute labellisés par P. Schaeffer :

- l'*écoute causale* : se renseigner sur la source - inclus les sentiments que peut éprouver le locuteur ;

- l'*écoute sémantique* : ne s'intéresser qu'au contenu verbal ;

- l'*écoute réduite* : goûter la musicalité de la voix, le plaisir d'entendre "l'inflexion des voix chères", l'informulé dans le parler...

Dans l'expérience quotidienne, les frontières ne sont pas aussi nettes. Par exemple, goûter le grain d'une voix (écoute réduite) mène presque fatalement à modéliser les sentiments, les désirs, la personnalité de qui parle (écoute causale), sans vraiment réussir à chasser totalement le contenu verbal (écoute sémantique). Barthes l'avait remarqué : le grain de la voix est un « mixte érotique de timbre et de langage », une « stéréophonie de la chair profonde », dans ces moments où « ça granule, ça grésille, ça caresse, ça râpe », ces moments où le signifié est « déporté très loin » (1993 : 1528-29).

L'écoute causale étant grandement gouvernée par les processus nerveux ascendants (courants *bottom-up*), il est hors de question de la débrancher ; l'*ancrage* des voix est donc une pratique courante. Dans le monde réel, trois niveaux d'analyse déterminent l'ancrage :

- le synchronisme, analyse temporelle (parole et spectacle de la source en train

d'émettre semblent être dans le même temps *vs.* décalés) ;

- l'espace, analyse de la couleur sonore, des pourcentages relatifs entre ondes directes et ondes réfléchies (parole et spectacle de la source en train d'émettre semblent être dans le même espace *vs.* des espaces différents) ;

- les cadres de vraisemblance (parole et source sont susceptibles d'appartenir au même script *vs.* des scripts différents ; un exemple de script est la présence possible d'un haut-parleur diffusant une voix avec une qualité mimétique suffisante pour en faire un double exact d'une voix *live*) ;

Au cinéma, on retrouve ces trois niveaux d'analyse, doublés par des scripts (processus descendants, ou *top-down*) relatifs aux usages techniques et esthétiques, et à leurs changements au cours de l'histoire du cinéma. Le synchronisme est analysé avec davantage d'indulgence par le spectateur courant, qui a intégré les pratiques du doublage. De même, l'analyse des composants du son est filtrée, chez le spectateur familier, par un savoir technologique (un son perçu comme « sourd » dans un film des années 90 passerait pour un son « standard » dans un film des années 40, à cause des progrès dans la bande-passante des supports d'enregistrement). Le troisième niveau d'analyse, enfin, réside essentiellement dans la question de l'attribution d'un monde de référence ; la voix peut appartenir (1) au monde diégétique (2) à un hétéro-univers (voix intérieure, voix déformée par une écoute pathologique…) (3) aux fosses (voix off du conférencier extradiégétique…) (4) au monde réel (indications de jeu du metteur en scène, « coupez ! » non effacé…).

À la TV, la situation est plus complexe, l'énonciation plus polyphonique (terme de Fr. Jost) et la stratification des mondes plus floue. Certains outils en provenance des études filmiques y fonctionnent mal - les axes d'opposition champ/hors champ, et diégétique/extra-diégétique y sont moins pertinents, par exemple. De surcroît, déterminations et consignes règlent devant le petit écran un usage *privé*, non *public* de la machine : il y a donc un champ des comportements autorisés beaucoup plus large.

La perception de la diégèse est plus complexe qu'au cinéma, ainsi le JT apparaît stratifié sur cinq niveaux simultanés :

- les événements du monde historicisés par le choix de la rédaction ;

- en filigrane, en arrière-plan diégétique, l'« autre partie du monde », celle qui n'est pas traitée ;

- le récit du JT lui-même, avec le générique d'ouverture, les rubriques... ;

- la partie métadiscursive de ce récit, beaucoup plus exhibée qu'au cinéma, justement grâce à la parole (« maintenant passons à... pour finir... ») ;

- en arrière-plan métadiscursif aussi, les « journaux parallèles » (guignols, vrai journal...) qui épinglent certains tics et rejaillissent sur la réception de leurs modèles-mêmes, en vertu du phénomène post-moderne qui veut que l'original se mette à ressembler à sa copie...

Pour être précis, événements du monde historicisés par le choix de la rédaction sont eux-mêmes gratifiés d'un degré de réalité qui fluctue énormément d'un spectateur à l'autre. Ce degré combine les trois types de croyance recensés par le chercheur américain R. Allen (1993) :

- *l'illusion reproductive* : modèle de la fenêtre ouverte sur le monde ;

- *l'illusion projective* : modèle de la simulation fictionnelle ; l'idée de reportage s'efface au profit de l'absorption diégétique, de la compassion, de l'identification secondaire, comme dans le cinéma classique ;

- la *lecture réaliste* : modèle de l'incroyance ; les événements qui se déroulent sous nos yeux ont lieu parce qu'une caméra était là pour les enregistrer, je ne vois ni le monde réel ni

une histoire fictive.

Fr. Jost a analysé du point de vue de l'énonciation ce foisonnement d'attitudes possibles, sous le nom de *feintise* (1999 : 32-34). Il est certain qu'à un certain niveau de stratification des instances énonciatives, le téléspectateur sera tenté de passer de l'attitude « être dedans » à l'attitude « être devant », ce qui lui permet d'ancrer toutes les voix, à égalité, dans la « boîte-source-réelle » qu'est le récepteur de TV. On peut penser que cette attitude est plus volontiers adoptée à des moments de grande stratification énonciative, par exemple à l'occasion des intermèdes où se mêlent annonces des programmes à venir, extraits de fictions et voix publicitaires, plutôt qu'à des moments qui rappellent le sage découpage diégétique opéré traditionnellement par le cinéma.

Ce tableau des attitudes d'écoute, aussi complexe qu'il soit, demande néanmoins à être encore affiné. Le téléspectateur, à mesure que le savoir relatif à la technique audiovisuelle diffuse dans l'espace public (par l'utilisation domestique du caméscope, du montage informatique à domicile… etc.), se présente en effet de plus en plus comme quelqu'un qui *expertise* la voix qui lui arrive en même temps qu'il l'ancre et décode ce qu'elle convoie.

TECHNOLOGIE DE LA VOIX TV

Les voix, comme tous les sons, véhiculent deux types d'informations : à propos de leur *source*, et à propos de leur *espace d'émission*. Les sons diffusés par le biais d'un haut-parleur, comme la parole TV donc, convoient en outre des informations (1) sur le dispositif technique de captage (ou de synthèse, puisque c'est la seule autre possibilité) (2) sur le dispositif technique de mixage/convoyage du signal (3) sur le dispositif technique de diffusion du signal (HP de la TV) (4) sur l'espace de diffusion (salon du téléspectateur). Reprenons en détail ces informations.

(1) Côté TV, captage.

Suivant les émissions, on oscille entre *l'exhibition* de la machine (les microphones) et sa *dissimulation*. Pour reprendre les distinctions de N. Burch, il y a *exhibition* lorsque l'émission est de type "confrontation exhibitionniste" (le microphone-orthèse renvoie à l'idée de l'*entertainer* - celui qui tient ensemble un groupe - se donnant en spectacle et s'adressant au large public d'une salle de spectacle), et *dissimulation* lorsque l'émission sollicite une absorption diégétique.

Mais dans les deux cas, on remarque que la prise de son est réglée de manière à cacher le « museau » (Barthes) : les scories de l'acte de parole sont éliminées (plosives postillonnantes, râles, souffles...). C'est une démarche de *lissage*, de type cosmétique, comme le maquillage des invités sur le plateau ou l'absence d'ombre sur ce même plateau, démarche qui se poursuit au mixage. Cette prise de son qui refuse la perche nie totalement les informations spatiales, ce qui renvoie à la « terreur » du hors-champ contigu décrite par S. Daney dans *Le salaire du zappeur* - c'est le fantasme télévisuel de la lucarne qui montre tout et ne cache rien. Le lissage est l'une des opérations qui motivent la remarque ironique de Godard (dans le CD Radio-France de ses entretiens avec Thierry Jousse) : « On entend mieux, oui, mais qu'est-ce qu'on entend ? »…

(1bis) Côté TV, synthèse.

Il faut bien entendu se garder de poser une dichotomie parole-trace *vs.* parole synthétique. Toute une gradation existe, comme pour l'image, avec en majorité des produits hybrides (*cf.* la question du mixage, ci-après). En l'état actuel de la recherche, il est *impossible* de calculer entièrement une voix synthétique de façon à tromper l'auditeur quant à son essence ; il y a toujours des échantillonnages dans ce qui passe aujourd'hui pour des "voix de

synthèse" - même chose pour le violon solo ou la figure humaine. Ceci est dû au chaos non déterministe qui régit le détail de ces objets, ainsi qu'à la grande compétence du système cognitif quant à leur analyse (on se doit de lire le moindre détail d'un visage, la plus minuscule inflexion de voix, pour survivre socialement). De surcroît les machines ne peuvent pas pratiquer l'écoute sémantique - pas plus qu'elles ne peuvent corriger les fautes de grammaire, car la grammaire fonctionne toujours (contrairement au vocabulaire) de manière contextuelle.

(2) Côté TV, mixage

Toutes les voix TV passent par des machines appelées *compresseurs*, qui poursuivent la démarche de lissage commencée à la captation. Le compresseur remonte en dynamique les syllabes ou mots un peu faibles, et écrase ceux qui sont émis de façon trop forte. D'autres machines, comme l'*égaliseur paramétrique* (*equalizer*), opérent un *tri* sur les fréquences émises, réalisent un « adoucissement » (*smoothing)*. Dans certaines émissions de variété, on pratique même ce qui est courant dans les radios commerciales, l'adjonction de *fréquences basses* - en tant que les sons graves, pour des raisons trop longues à exposer ici, sont privilégiés dans le contexte postmoderne.

Les informations spatiales – évacuées, on l'a vu, de la prise de son - sont éventuellement recréées ici, au mixage (je soupçonne l'ingénieur du son de la messe en direct, le dimanche matin, de reconstituer la réverbération de type église - rien de plus facile, il suffit d'appuyer sur un bouton, et cela permet d'évacuer les accès de toux et autres chaises qui grincent, bruits qui seraient enregistrés par une prise de son naturellement spatialisée), ou tout bonnement créées ex nihilo : une petite queue de réverbération est généralement perçue comme agréable, *enveloppant*, par l'auditeur.

Dans certaines émissions des chaînes musicales, il peut aussi y avoir musicalisation : la voix de la présentatrice de l'émission *Amour*, sur MTV à la fin des années 90, était probablement gratifiée d'un soupçon de *flanger* ou d'un quelconque *harmonizer* de façon à exemplifier le grain du violoncelle, sinon, par un « transfert figuratif » basé sur l'intermodalité audio-haptique déjà signalée, celui du velours…

Il résulte de tout ce travail masquant un phénomène de *déréalisation invisibilisée*, qui donne lieu, dans sa modeste part, au phénomène d'hyper-réalité décrit par J. Baudrillard : le téléspectateur a tendance à croire que les présentateurs TV ont cette voix chaude régulière et veloutée dans la vie réelle... L'*iconolâtrie* (croyance en la vérité objective de l'image-trace) est par ailleurs encore plus flagrante dans le domaine sonore ; peut-être doit-on parler d'*acousmolâtrie*...

À l'inverse, le mixage peut *s'exhiber* - en général quand l'émission est du domaine de l'*œuvre*, et non pas comme le JT du domaine de la *transmission*. À notre époque de haute-fidélité des systèmes de reproduction acoustique, une baisse brutale du degré d'iconicité sonore, par exemple, lorsqu'elle est perçue comme délibérée par le téléspectateur, est susceptible de faire sens. La voix enregistrée salement (*lo-fi*) connotera ainsi le dédain pour l'univers *high-tech* au profit du côté familial "rétro" de l'image-trace (le petit dernier qu'on faisait parler devant le magnétophone…). etc.). Dans le même ordre d'idées, dans certaines émissions (*Taratata*) on observe un *décalage* d'un ou deux photogrammes/seconde dans le synchronisme audiovisuel. Cette manipulation, en général associée à une granulation artificielle de l'image (via un logiciel comme *Grain management*), est destinée à singer - pour s'en approprier le prestige vacillant - le glorieux ancêtre *cinéma*, avec ses imperfections de machine 19ème... Fr. Jost y verrait sans nul doute une autre forme de *feintise*…

(2bis) Côté TV, convoyage du signal.

Il est à noter qu'en cas de problème de diffusion du signal (décrochages, interférences...) les sautes sont audiovisuellement synchrones, ce qui n'est pas le cas au cinéma. Ce synchronisme asseoit l'idée de *causalité verticale* (croyance selon laquelle son &

image sont, comme dans la vie, toujours liés causalement), et conforte le spectateur dans l'« acousmolâtrie ».

(3/4) Côté téléspectateur : HP de la TV.

On retrouve l'oscillation entre l'exhibition de la machine et la dissimulation illusionniste du "truc". Certains modèles de TV Sony sont par exemple proposés exactement au même prix avec HP apparents en façade ou cachés sur les flancs à la faveur du rétrécissement du tube… Un axe d'opposition se dessine ainsi entre deux dispositifs extrêmes :

- le *dispositif de ventriloquie* qui rejoint le "cinéma qui triche" (M. Duras) et induit l'illusionnisme, l'absorption diégétique, le "il me parle" ;

- l'*exemplification technologique,* comme dans les salles de cinéma *high-tech*, le "ça parle".

Il va de soi que le premier dispositif convient mieux lorsque le spectateur décide d'« être dedans » et le second lorsqu'il préfère « être devant »,

TRIBUNE/ECRAN : LE POUVOIR DE LA VOIX

Les cogniticiens ont une conception *constructiviste* de la perception – l'homme voit moins *ce qui existe* que ce qu'il *croit* pouvoir exister. Différentes composantes du dispositif TV – en premier lieu la captation et le mixage – vont pré-régler pour lui les relations audiovisuelles dans un certain sens. Captage et mixage des voix privilégieront l'idée de causalité verticale, donc de *conception soustractive* du médium (croyance bazinienne selon laquelle le médium *rend* le monde *moins* quelques détails comme le relief, les odeurs…). Associée à l'idée de direct et de flux ininterrompu, la conception soustractive mène au sentiment d'*hyperréalité*. Cette lecture se verra facilitée par les *modèles domestiques* et leurs TCN associées, tels qu'ils sont intégrés par une majorité de téléspectateurs (TCN = Théories Causales Naïves ; la causalité verticale en est une). Une question demeure cependant en suspens : que se passe-t-il avec des émissions qui exhibent une scission point de vue/point d'écoute, empêchant de prime abord la construction soustractive, que se passe-t-il, donc, avec la *Tribune* ?

Pour reprendre la distinction de G. Soulez *Tribune/Ecran*, l'Ecran, fonctionne plutôt sur le modèle de la prise d'empreintes synchrones, comme dans le cinéma classique (il "neutralise l'axe de la parole de l'orateur, pour considérer l'arrangement des formes entre elles"). La Tribune (qui "organise l'apparition des formes autour d'un axe qui est la parole de l'orateur"), et la Tribune virtuelle ("orateur invisible"), sont des concepts plus délicats. Il est peu probable que la Tribune virtuelle mette en péril la conception soustractive du médium. Elle exemplifie en effet le modèle le plus ancien, dans les arts audio-visuels, celui de la *conférence* (lanterne magique) Comme au cinéma, le spectateur matérialise une *fosse* plus ou moins proche de lui ou, à l'inverse, proche de l'écran, selon que le commentateur off donne l'impression de découvrir le film en même temps que lui (= d'être un co-spectateur) ou alors de fabriquer le film par l'omnipotence de sa voix (narration déléguée). À la TV le système va être plus emboîté (*chinese-boxed*), car le système des commentateurs peut se complexifier, par exemple :

- le plateau, qui se décompose en présentateur dans le champ et mystérieuse régie hors-champ ;

- le direct-ailleurs, par exemple un stade, avec son propre commentateur hors-champ et son événement dans le champ.

Le téléspectateur est laissé très libre en tant que les connexions spatiales sont laissées à sa convenance : par exemple en cas de cafouillage en régie, le présentateur va

regarder dans le vague pour bafouiller... on ne saura pas où est le technicien qui s'est trompé de bande : à droite, à gauche, au-dessus... L'une des conséquences de ce flou est que l'orateur d'une Tribune (non virtuelle) peut tout de même exercer un grand pouvoir sur la lecture de l'image par le spectateur : l'évanescence du véritable espace de pouvoir (la régie) l'autorise même souvent à *incarner* l'énonciation - ainsi au JT, dont la partie métadiscursive est justement exhibée grâce à la parole. Seules des pratiques marginales d'*occurrences vides* (sur le générique de fin, comme chez P. Amar sur Paris Première) ou de *fade in/fade out* un peu trop saillants (pour cause de prises de parole simultanées sur le plateau de *Droit de réponse*), connotent l'idée d'un travail caché et une *conception additive* du médium, comme dans le cinéma de la Modernité.

Les modèles domestiques disponibles dans l'espace public des sociétés industrialisées privilégient de surcroît la supériorité de la voix sur l'image au sein du modèle de la Tribune. Il faut en effet rappeler qu'à l'inverse de ce qui s'est passé pour le cinéma, où le son solidaire est arrivé trente ans après l'image, la télévision s'est répandue massivement *après* la radio, et la vague du numérique domestique a commencé par le son (platine CD) *avant* l'image (le vidéodisque, le DVD et les installations *home cinema* sont venues par la suite). Autres grosses différences confortant la hiérarchie dans la Tribune (1) il n'y a quasiment pas de *pratique familiale* de la prise de sons en tant que telle, passée une vogue des magnétophones à bande dans les années soixante, alors qu'il y a une énorme pratique familiale des images (2) l'appareil audiovisuel dominant, le caméscope, est une machine anthropomorphe littéralement construite autour de l'idée de causalité verticale en perspective naturaliste (= tous les sons près de la caméra sont enregistrés ; le reste, non).De cela il résulte qu'un amateur muni d'un caméscope DV bas de gamme peut faire de belles images qui pourront être diffusées sans problème au JT, tandis qu'en revanche la réussite d'une bande-son reste un énorme travail difficile à réussir seul. Les sons, et surtout les voix off, vont ainsi garder une part de *mystère* propice à l'épanouissement de leur pouvoir de *guides*, de *schémas de construction* de l'image par celui qui la voit. La combinaison des deux modèles archaïques, celui de la mère (histoire familiale/mémoire épisodique) et celui du conférencier-lanterniste (Histoire avec un H/mémoire sémantique) ne peut dans ces conditions qu'aboutir à des figures d'*éclaireurs*, énonciateurs-professeurs virtuels et démiurgiques susurrant avec plus ou moins de succès « voyez ceci, je le veux ».

On ne surestimera pas, toutefois, le pouvoir de ces voix. D'abord parce que c'est le langage lui-même, et l'acte même de prendre la parole, qui conditionnent la réussite de son exercice. « Rien à faire : le langage, c'est toujours de la puissance ; parler, c'est exercer une volonté de pouvoir : dans l'espace de la parole, aucune innocence, aucune sécurité », écrit Barthes (1993 : 1195). Ensuite parce qu'avant d'appeler une écoute sémantique conforme à une intention dénominative et déictique (« ici, à gauche de l'image, c'est… là, à droite, vous voyez… »), la voix déclenche de façon quasi pavlovienne une écoute causale visant à détecter les émotions du locuteur. Pour poursuivre avec Barthes, la « parole publique » mène « tout droit au divan » : la « barbe postiche » de celui qui parle, comme dans un cauchemar, « se décolle par lambeaux devant tout le monde » (*op. cit.* pp. 1196-97). Avant de guider la lecture, le locuteur nous renseigne donc sur lui-même, ou sur un alter ego de lui-même que nous modélisons en fonction du grain et des inflexions de sa voix – « il n'y a pas de voix neutre – et si parfois ce neutre, ce blanc de la voix advient, c'est pour nous une grande terreur, comme si nous découvrions avec effroi un monde figé, où le désir serait mort » (Barthes 1994 : 881). Lissage, compression et polyphonie énonciative conduisant parfois par dépit ou lassitude à *être devant* sont autant de gouttes de colle visant à fixer cette barbe postiche, mais bien rares sont les cas de vraie neutralité ou de barbe si convaincante qu'elle passe pour être vraie. Enfin, pour citer le dernier terme de la trichotomie des écoutes de Schaeffer, le pouvoir de guide est souvent concurrencé, chez l'auditeur, par la tentation de l'écoute réduite, surtout si la voix est musicalisée par un mixage destiné à flatter l'oreille en gommant les traces du museau.

BIBLIOGRAPHIE

Allen Richard 1993 : « Representation, illusion, and the cinema », *Cinema Journal* n°32 vol. 2.

Bailblé Claude 1999 : *La perception et l'attention modifiées par le dispositif cinéma*, Thèse de doctorat en esthétique, Edmond Couchot dir., Paris VIII.

Barthes Roland 1993, 1994, 1995 : *Oeuvres complètes tome I 1942-1965*, *tome II 1966-1973*, *tome III 1974-1980*, Eric Marty éd., Le Seuil, Paris.

Burch Noël : *La lucarne de l'infini*, Nathan, Paris 1991 (trad. fr. de *Life to those shadows*, Univ. of California Press 1990).

Jost François 1999 : *Introduction à l'analyse de la TV*, Ellipses, Paris

Kozloff Sarah 1988 : *Invisible storytellers (Voice-over narration in american fiction film)*, The university of California Press, Berkeley / Los Angeles / Londres.

Schaeffer Pierre1977 : *Traité des objets musicaux*, Seuil, 3ème éd. augmentée, Paris (1ère éd. 1966).

Pour citer ce texte : L. Jullier, « Etre devant/Etre dedans. Son & parole à la télévision d'un point de vue cognitiviste », intervention à Téléparoles, séminaire INA/Univ. de Metz, G. Soulez dir., 2000.