UNIVERSITE DE TOULOUSE LE MIRAIL

LA.R.A

ANALYSE ACOUSTIQUE de la VOIX

pour la

DÉTECTION de PERTURBATIONS

PSYCHOPHYSIOLOGIQUES

APPLICATION au CONTEXTE AÉRONAUTIQUE

Robert Ruiz

Laboratoire de Recherche en Audiovisuel (LA.R.A) - Université de Toulouse 2 – le Mirail

SYNTHÈSE des TRAVAUX

Mars 2012

AVANT -PROPOS

Ce manuscrit est la partie principale du mémoire présenté pour l'habilitation à diriger des recherches de l'auteur.

Il présente une revue générale des connaissances sur le thème des relations entre la voix et les perturbations psychophysiologiques du locuteur puis une synthèse des travaux de recherche. Ils traitent de l'analyse acoustique de la voix appliquée aux perturbations spécifiques que rencontrent les personnels navigants.

Le Bureau d'Enquêtes et d'Analyses (B.E.A) pour la sécurité de l'aviation civile est l'initiateur et le soutien de ces études. Depuis plus de dix ans il renouvelle sa confiance. Ce sont donc de très chaleureux remerciements qui lui sont adressés et en particulier à Philippe Plantin de Hugues, l'interlocuteur du BEA. Les travaux doivent beaucoup à ses initiatives, à son écoute attentive, à ses conseils avisés, le tout dosé dans un subtil mélange laissant au chercheur une grande autonomie pour ses activités.

1. INTRODUCTION : CADRE DE LA RECHERCHE

La parole, faculté de parler, renvoie à la notion de communication. Elle utilise un langage qui code l'information à transmettre par un ensemble de signes. Ceux-ci ne sont pas exclusivement vocaux mais peuvent être gestuels, tactiles en ayant recours à des attitudes corporelles spécifiques : regards, larmes, rires etc.

La parole est l'élément sonore du langage. Elle permet de communiquer oralement avec le code sémantique et syntaxique de la langue.

Une émotion peut être exprimée en choisissant les éléments lexicaux adaptés et la syntaxe appropriée. Si tel était le seul moyen d'y parvenir, l'expression verbale ne diffèrerait de l'écriture que parce qu'elle est sonore. Mais la parole est l'émanation d'un être vivant. Le corps et la pensée vont non seulement personnaliser sa parole et la singulariser de toutes les autres, mais également témoigner, consciemment ou non, des émotions, des sentiments, des aspirations, des besoins, de la posture, de sensations d'affaiblissement physique ou de plus grande activité, de l'état de l'appareil phonatoire mais aussi de l'état général de l'organisme.

La communication verbale utilise des mécanismes du langage qu'étudie et décrit la linguistique dans ses dimensions stylistiques, phonétiques, phonologiques, historiques, sociologiques, psycholinguistiques etc. Elle possède également une composante para-linguistique comprenant les moyens de communication qui accompagnent et renforcent la parole comme les expressions du visage, les gestes. Elle porte enfin une composante non linguistique chargée des éléments psychophysiologiques précédents et que l'acoustique de la voix peut identifier.

La parole est donc le siège d'expressions verbales non linguistiques qui traduisent l'état psychologique et physiologique de l'individu.

L'ensemble des sons qu'il produit, sa voix, sont influencés par ces facteurs personnels. L'influence peut d'ailleurs être perçue, ressentie par l'auditeur et conduire à modifier sa réaction.

L'acousticien peut enregistrer la communication et étudier les signaux électriques qui en résultent. Deux approches existent pour analyser la voix.

La première consiste à identifier les propriétés acoustiques des éléments du code de la langue : consonnes, voyelles, transitions entre consonnes et voyelles etc. Elle en examine également la variabilité intra - et inter - locuteur. C'est ainsi que chaque type de voyelle a pu être caractérisé par des valeurs numériques de paramètres acoustiques assorties d'intervalles de confiance. La connaissance en la matière est aujourd'hui importante (R.Boite 1987, Calliope 1989 au §5.2.5 page 142 du manuscrit).

La seconde s'attache à isoler les signes acoustiques des expressions psychophysiologiques. Initialement, elle n'avait sans doute d'autre objet que le progrès de la connaissance scientifique mais rapidement les applications possibles ont soulevé des intérêts notamment en matière de sécurité. Par exemple, il pourrait s'agir d'identifier une voix mensongère ou de détecter un trouble lors d'une reconnaissance vocale automatique du locuteur ou encore d'estimer un stress.

Une autre application relève de la **sécurité des transports**. Les progrès technologiques sont tels que les facteurs humains sont devenus une cause prépondérante des accidents.

Ce sont des facteurs qui affectent les performances : l'état de forme, la santé, la pression des horaires, la surcharge et la sous-charge de travail, le sommeil et la fatigue, l'alcool, les médicaments et les produits psychotropes. Ce sont aussi des éléments psychologiques (état d'anxiété, confiance, peur, stress, gestion des risques …) et psychosociaux (motivation, travail en équipe, responsabilité individuelle et de groupe). Ils peuvent aussi être multifactoriels.

Ces facteurs humains sont à plus de 75% la cause des accidents aéronautiques loin devant les causes techniques. Il est donc devenu indispensable de s'interroger sur le comportement humain sous diverses formes comme pour diverses cultures. La problématique est rendue plus complexe par l'augmentation du traffic aérien (5% par an environ). À ce rythme, on peut atteindre plusieurs accidents majeurs par mois sans que le taux d'accident mondial augmente. En matière d'accidentologie des transports, l'impact du nombre absolu d'accidents est énorme pour le grand public et les médias. Il est donc capital de comprendre et d'agir sur le comportement humain en situation de pilotage pour réduire le nombre d'accidents.

Les sciences cognitives, l'ergonomie (en particulier le travail sur l'interface homme-machine), les actions de formation, les apports technologiques y contribuent. L'approche acoustique participe elle aussi à cette contribution multidisciplinaire. Elle ne s'inscrit pas dans le champ de la compréhension du rôle des facteurs humains mais dans celui de la mise en œuvre de moyens de détection et d'analyse pour agir sur le comportement et prévenir un possible accident.

Il importe pour cela de disposer d'une appréciation de l'état psychophysiologique de l'individu. L'emploi d'indicateurs physiologiques biochimiques, cardiologiques, de l'action motrice (temps de réaction, mobilité oculaire) ou autres, comme la température et la réponse galvanique de la peau, l'électro-encéphalographie est possible mais les paramètres acoustiques de la voix semblent tout particulièrement adaptés. La prise de son est non invasive surtout pour une population de locuteurs dont les métiers nécessitent de communiquer via une chaîne électro-acoustique (pilotes d'avion, contrôleurs du ciel).

Tel est l'intérêt de l'approche acoustique. L'étude bibliographique développée dans le manuscrit montrera qu'il n'est pas vain de chercher dans cette direction un moyen de parvenir à estimer l'état psychophysiologique d'un locuteur. La voix se modifie sous l'effet des facteurs qui affectent la performance. Mais il n'existe pas de paramètres acoustiques, ou de combinaison de ceux-ci, dont les mesures puissent attester de cet état. La relation entre les facteurs humains et leurs conséquences vocales n'est pas bijective. On mesure les implications des premiers mais l'analyse d'une voix « perturbée » ne permet pas encore d'identifier le type de facteur qui en est à l'origine ni le degré de trouble du locuteur. Même avec la connaissance d'informations contextuelles sur l'origine de l'enregistrement cela n'est pas possible. Concrètement, l'analyse acoustique de l'enregistreur phonique d'un avion accidenté dont on connaît les paramètres de vol ne permet pas de statuer sur une cause humaine liée à la fatigue ou à une perte de vigilance ou à un stress etc … La voix est le siège de nombreuses et diverses manifestations émotionnelles et physiologiques. La recherche de leur trace acoustique propre n'a pas pleinement abouti. La possibilité d'y parvenir existe.

Parmi l'ensemble des facteurs humains agissant sur le comportement et parmi les nombreuses applications de la recherche de leurs conséquences, l'étude se restreint ici au seul secteur aéronautique et aux facteurs qui concernent spécifiquement le domaine.

C'est pourquoi, la grande majorité des expérimentations mises en œuvre sont rattachées au secteur aéronautique.

D'autres voies de recherche ont été explorées, comme l'interaction entre le contexte émotionnel de l'image et l'expression orale ou encore, l'effet d'interférence de la lecture d'un mot sur la dénomination de sa couleur.

Si cette recherche trouve sa cohérence dans les applications principales auxquelles elle se destine, les mesures acoustiques peuvent être confrontées à d'autres contextes expérimentaux.

La méthode expérimentale a associé des expériences de laboratoire à l'analyse de cas réels : le dépouillement de l'enregistreur phonique d'un avion accidenté et la simulation d'incidents en simulateur. De même, il s'est agit d'enregistrer des pilotes au cours d'une journée de vol et de conduire une expérience similaire au cours d'une journée de conduite automobile.

Les expériences entreprises couvrent le champ des perturbations psychophysiologiques que peuvent rencontrer les acteurs aux commandes du transport aérien.

En premier lieu, ce sont des syndromes de l'adaptation à des situations inattendues et graves, ou pouvant le devenir. C'est l'émotion et le stress que l'apparition ou la gestion d'un dysfonctionnement font naître.

En second lieu, il s'agit des effets de la charge et des conditions de travail. La succession des vols court-courriers pendant plusieurs jours, dans un état de concentration et de vigilance importantes du pilote, peut induire de la fatigue. De la somnolence due à une privation de sommeil peut également apparaître compte tenu de l'heure matinale des premiers vols. À l'opposé, lors de vols long-courriers, une capacité à promptement réagir est nécessaire si un incident se manifeste alors que le pilote se repose.

Ces situations ont été individuellement étudiées.

L'orientation de la recherche a été d'examiner l'ensemble du champ des perturbations psychophysiologiques du contexte aéronautique plutôt que de multiplier les expérimentations pour l'une d'entre elles. Elles sont d'une mise en œuvre parfois difficile et la grande quantité d'information à traiter a conduit à restreindre le nombre de locuteurs. Ces travaux revêtent donc un caractère exploratoire mais global qui les dispose désormais à être élargis.

La méthode d'analyse des mesures acoustiques a consisté à établir un état de la variabilité naturelle du locuteur pour chaque paramètre acoustique afin de lui comparer les prononciations susceptibles d'être modifiées par les diverses causes perturbatrices. La

comparaison consiste à observer le contour temporel des grandeurs mesurées sur des tracés d'évolution chronologique et à effectuer des tests statistiques de comparaison de moyennes.

Souvent, le nombre de phonèmes prononcés est faible lorsque l'effet présumé des perturbations se manifeste et ce, qu'il s'agisse de vocabulaire contrôlé lors des expériences de laboratoire ou de discours naturel. L'analyse statistique compare alors un échantillon de petite taille, avec un échantillon plus important relatif à l'état de « repos » du locuteur. C'est pourquoi, la méthode statistique peut conduire à conclure à des variations non significatives alors que l'examen des tracés chronologiques accompagné de calculs de régressions permet de noter des tendances.

Le travail d'analyse statistique reconnaît implicitement que les signes acoustiques des perturbations psychophysiologiques se situent au delà des limites supérieure et inférieure de la variabilité normale à l'état de repos ou dans un état de neutralité émotionnelle du locuteur. C'est une hypothèse, mais il est possible qu'ils n'y soient pas inclus. D'où la grande importance accordée à l'examen des variations chronologiques des paramètres acoustiques.

L'information transmise par le signal vocal est conséquente et des choix ont été faits pour en restreindre la quantité et la nature.

Les essais entrepris avec le signal de certaines consonnes ont conduit à estimer que la variabilité des mesures était telle, qu'elle pouvait masquer les effets recherchés de la perturbation du locuteur. En conséquence, les signaux traités sont ceux des voyelles.

Cette restriction peut être utile pour des applications futures. La segmentation des voyelles, telle qu'elle a été réalisée, est moins sujette à des erreurs de frontière eut égard à la brièveté de la plupart des consonnes. Ces dernières peuvent nécessiter de numériser le signal à des fréquences d'échantillonnage incompatibles avec les systèmes d'enregistrement embarqués. Enfin, il a été décidé de ne conserver que la partie quasi – stationnaire du signal de voyelle et d'éliminer l'attaque et l'extinction. Ce choix rend compatible une future extraction automatique des segments de phonèmes en temps quasi – réel.

La contribution est double. Premièrement, elle tient au fait même des situations expérimentales examinées et à la globalité du champ traité.

Deuxièmement, elle résulte de la mesure d'un nombre varié de paramètres acoustiques, même si ces caractéristiques n'ont pas toutes été mesurées dans toutes les expérimentations pour cause de progression dans la recherche et dans les moyens d'analyse et de calcul.

Elle est particulièrement associée à la proposition de trois paramètres originaux. Ces derniers ne sont pas toujours en capacité à détecter une perturbation psychophysiologique du locuteur. Mais ils peuvent compléter la liste des mesures possibles sur le signal vocal et être confrontés à d'autres situations expérimentales. Il s'agit :

- de la fréquence d'équilibre spectral qui est la plus petite distance entre une distribution cumulée basse fréquence et une distribution cumulée haute fréquence des niveaux sonores dans le spectre ;
- du « Digital Amplitude Length » (D.A.L) et de ses paramètres dérivés qui renseignent sur l'amplitude parcourue par le signal au cours d'une tranche temporelle ;
- d'une fréquence de modulation d'amplitude des maxima de niveau sonore.

Enfin, une méthode de modélisation a été introduite. Elle propose de fonder la modélisation acoustique sur les mesures d'autres types de paramètres comme les caractéristiques physiologiques.

Pour les recherches sur l'intelligibilité de la parole en milieu bruyant, des tests de jurys ont permis d'établir une évolution « en S » du nombre d'éléments de vocabulaire reconnus en fonction du rapport signal sur bruit. Des modèles acoustiques intégrant des mesures issues de la réponse impulsionnelle du local ont alors pu être développés pour suivre le même type d'évolution en fonction du rapport signal sur bruit.

Ici, la méthode peut être identique en s'appuyant sur des modèles issus de la mesure de paramètres physiologiques ou de résultats d'enquête. Cette approche a été rendue possible par l'existence de travaux de ce type qui proposent une variation linéaire chronologique de la fatigue de pilotes. C'est pourquoi une modélisation acoustique multi – paramètres de variation linéaire chronologique est introduite et testée pour les expérimentations sur la fatigue.

Dans la première partie qui va suivre, une revue générale des connaissances est proposée.

Elle cible en premier lieu les travaux réalisés pour l'analyse de la voix dans le secteur des transports aux fins de détecter des perturbations psychophysiologiques.

Viennent ensuite la synthèse des recherches établissant des liens entre la voix et l'émotion et la voix et la fatigue. Ce sont en effet les deux principaux facteurs perturbateurs qui se présentent dans les contextes aéronautiques.

Le travail bibliographique est ensuite complété par l'examen d'autres causes possibles de modifications de la voix comme le mensonge ou la consommation d'alcool.

Des pathologies générales d'origine psychologique ou physiologique provoquent également des altérations du signal vocal. Dans la synthèse des connaissances ne figurent pas celles qui traitent de pathologies de l'appareil phonatoire car elles sortent du cadre des recherches entreprises. La problématique est différente et les locuteurs du secteur des transports ne sont pas concernés par ces aspects dans le cadre de leur activité professionnelle.

De même, hors du cadre de cette recherche, les travaux relatifs à la qualité de la voix (claire, chuchotée, grinçante, rauque …) ne sont pas directement mentionnés.

Les références bibliographiques qui complètent le manuscrit sont les publications de l'auteur puis celles relatives aux principaux travaux internationaux sans pour autant que la liste proposée soit exhaustive.

Suite à la revue générale des connaissances et à la présentation de la contribution, une synthèse et un examen critique des travaux de recherche concluent le document en s'attachant également à décrire l'orientation des travaux en cours.

2. REVUE GÉNÉRALE DES CONNAISSANCES

Les travaux de recherche relatifs aux manifestations vocales de perturbations psychophysiologiques ne s'inscrivent pas tous dans le contexte des transports.

À titre d'exemple, des expérimentations ont eu lieu pour détecter et simuler des émotions par la voix, pour détecter la fatigue, le mensonge, pour mesurer les effets de la réalisation d'une tâche arithmétique ou d'exercices mentaux et psychomoteurs.

Un des propos de cette revue des connaissances est d'organiser thématiquement ces diverses expériences du point de vue de l'analyse acoustique du signal. Une approche linguistique pourrait également être envisagée comme l'illustre l'exemple suivant.

Considérons le cas du célèbre récit du journaliste radio, Herbert Morrison, témoin en direct de l'approche, puis de l'embrasement soudain du plus grand aéronef jamais construit, le Zeppelin « Hindenbourg » à Lakehurst (New Jersey, U.S.A) le 6 Mai 1937 :

« I…uh…can't even talk to people. It's a, it's a, it's a … I, a, I can't talk, ladies and gentlemen … I, I can hardly breathe. I, I'm going to step inside where I cannot see it. I, I can't a … Listen folks, I, I'm going to have a stop for a minute because I've lost my voice. This is the worst thing I've ever witnesed » (C.E.Williams, 1981).

L'importante résonance médiatique de l'accident et la création de deux commissions d'enquêtes allemmande et américaine ont sans doute contribué plus tard à la naissance de services spécifiques d'enquêtes et d'analyses dans les pays constructeurs de matériel de transport.

La transcription précédente permet de noter des signes syntaxiques de l'émotion et d'en ressentir les signes acoustiques. Il est donc certainement possible d'identifier la trace d'un facteur perturbateur du locuteur dans la transcription de ses propos. L'organisation des phrases, la nature du vocabulaire employé peuvent être des indicateurs du trouble. Cette approche n'est pas développée car elle n'entre pas dans la spécialité acoustique des travaux effectués.

Nous sommes en présence d'un cas rare et avéré de perturbation émotionnelle du locuteur otant le doute sur l'existence des modifications acoustiques à l'appui des impressions subjectives de l'écoute et de la syntaxe utilisée.

Tel qu'il a été présenté dans l'introduction précédente, le cadre de la recherche concerne principalement le domaine des transports.

C'est pourquoi la première partie de la revue des connaissances va s'attacher à décrire les travaux d'analyse vocale du secteur pour estimer les modifications de la voix sous l'effet des facteurs perturbateurs qui lui sont spécifiques (§ 2.1 page 8).

Le thème qui suit traite, hors du cadre des transports, des relations entre la voix et l'émotion non seulement sous l'angle de l'analyse mais aussi du point de vue de la synthèse sonore et de la reconnaissance automatique (§ 2.2 page 15). Comme on le verra, c'est un large champ d'investigation au cœur de la problématique de cette recherche.

Vient ensuite l'examen bibliographique des travaux sur la voix et la fatigue (§ 2.3 page 41).

La dernière partie couvre des champs de perturbations vocales très bien identifiés que sont le mensonge, les pathologies non vocales et l'alcoolémie (§ 2.4 page 43). Bien qu'ils

n'entrent pas directement dans l'application des modifications vocales rencontrées dans le contexte aéronautique, il est utile d'en faire état.

La détection de mensonge est pour l'essentiel fondée sur la mesure de micro-tremblements des cordes vocales qui mérite d'être appliquée aux situations expérimentales étudiées ici. Les variations de leur fréquence de vibration n'est peut-être pas exclusive à la manifestation du mensonge. Qui plus est, elles trouvent leur origine dans une étude sur la fatigue (W.A.Bousfield 1932).

Une synthèse générale des travaux (§ 2.5 page 46) conclura l'exposé des connaissances.

2.1. Analyse de la Voix dans le Secteur des Transports (références bibliographiques § 5.2.1. page 124)

Les travaux qui sont répertoriés dans ce paragraphe relèvent d'études spécifiques au domaine des transports et concernent des perturbations d'origine diverses. Le propos est de recenser et de regrouper ces études car elles entrent dans le même champ que les recherches entreprises et co-encadrées. Elles ne se subdivisent donc pas selon la perturbation mais selon le type de transport.

Transport Aérien et Spatial

● **Dans le secteur spatial**, la préoccupation majeure est d'évaluer à distance, par transmission radio, l'état psychophysiologique du spationaute. Il est donc à la fois question de stress (L.Khachatur'yants, 1972), de fatigue, de parole après le sommeil (V.Kuznetsov, 1975) que ce soit dans le véhicule spatial, lors de sorties extra – véhiculaires dans l'espace ou au cours de différentes phases de vol (P.V.Simonov, 1973 et 1977).

De nouveaux paramètres acoustiques sont introduits pour déterminer l'état physique et émotionnel du locuteur (P.V.Simonov, 1973). Selon les auteurs, ils ont été testés avec des acteurs de théâtre simulant des émotions mais aussi au cours de vols aériens et spatiaux.

Le premier, noté M augmente avec le stress émotionnel. Dans la gamme de fréquences du premier formant (300 – 1200 Hz), la puissance moyenne de sortie de plusieurs filtres est mesurée. M est la moyenne géométrique de ces puissances multipliée par un coefficient de proportionnalité égal à 1000 et par le rapport de leur somme à la puissance totale du signal mesurée avec un filtre linéaire. Des mesures du rythme cardiaque montrent leur corrélation avec les changements vocaux mesurés par le paramètre M.

Sont proposés ensuite plusieurs autres grandeurs qui ont pour rôle d'estimer le degré et la valence du stress émotionnel (émotions positives ou négatives). Le signal de parole est filtré par un banc de cinq filtres d'octave sur la durée de prononciation. Les cinq niveaux sonores qui en résultent sont alors combinés dans diverses formules les multipliant et les divisant entre eux pour obtenir cinq autres quantités. Des combinaisons linéaires de leurs valeurs permettent de différentier les émotions négatives des positives et d'en évaluer le degré.

Si la recherche présente des paramètres très originaux et semble-t-il capable de réaliser non seulement la détection de l'émotion mais aussi son degré et son type, les informations relatives aux filtrages effectués tout comme l'origine des formules proposées pour définir et combiner les paramètres ainsi que le détail des expérimentations manquent. C'est

certainement ce qui a conduit les recherches ultérieures à ne pas tester et confronter ces caractéristiques aux multiples contextes expérimentaux mis en œuvre.

Deux missions spatiales « Skylab » ont donné l'opportunité d'étudier le stress psychologique des astronautes suite à l'apparition de problèmes d'équipement à bord, de la réalisation de tâches complexes et difficiles et d'observer l'effet de la charge de travail en fonction du nombre d'heures d'activité (H.J.Older, 1975). L'objectif était de vérifier si un appareil spécifique, le « Psychological Stress Evaluator » (P.S.E) (cf §2.4.1 page 43), pouvait estimer le stress par analyse vocale et donc envisager que la NASA l'utilise pour de futures missions. Les conclusions n'ont pas recommandé l'utilisation du P.S.E comme indicateur du caractère stressant des tâches effectuées lors des missions spatiales en raison de l'absence de corrélation avérée entre les causes psychophysiologiques et les indications de l'appareil de mesure.

● **Pour l'aéronautique** la problématique est similaire à celle du secteur spatial (V.B.Maxwell, 1984) mais avec un volume du trafic bien supérieur et des enjeux de sécurité plus importants compte-tenu du nombre de passagers transportés. Aujourd'hui, la fiabilité des appareils est telle que les facteurs humains sont devenus une des causes majeures des accidents. Il peut s'agir par exemple de problèmes de communication ou de coordination des tâches entre les pilotes (M.Neville, 2005) dus aux facteurs perturbants que sont le stress (Hansen, 2000) ou la fatigue (J.Armentrout, 2006). Pour cette dernière cause, 186 accidents aéronautiques militaires ont été répertoriés comme liés ou dus à la fatigue entre 1971 et 1977 (G.P.Krueger, 1978).

Parallèlement aux études sur les vols réels, des travaux en simulateurs peuvent être réalisés car les pilotes sont formés sur ces outils reproduisant les conditions de vol de manière très réaliste.

Pour toutes ces raisons, le transport aérien civil et militaire (J.H.L.Hansen, 2000) nécessite et offre des moyens d'estimation de l'état psychophysiologique du pilote. L'analyse vocale est en premier lieu concernée parce que les communications sont orales et que l'on dispose, en cas d'accident, de l'enregistrement des conversations. Les études peuvent se subdiviser en deux catégories : l'étude de cas réels et celle en conditions simulées.

▪ ***Cas réels*** : Ce sont des accidents aéronautiques qui ont donné lieu à l'analyse de l'enregistreur des conversations ou à l'exploitation d'enregistrements des communications.

Les études sont peu nombreuses, les causes sont propres à chaque vol et dans la plupart des cas ne sont pas communiquées. Aussi, il est très difficile d'en dresser une typologie. On opte ici pour un inventaire des caractéristiques acoustiques que les auteurs ont décrit comme sensiblement modifiées par le stress ou l'émotion des pilotes.

Un paramètre original a été introduit dans les recherches, le VSSR ou « Vibration Space Shift Rate » (I.Kuroda, 1975). Il a été testé sur les enregistrements des communications vocales de pilotes entre l'avion et le sol impliqués dans 14 accidents dont 8 leur ont été fatals. Il prend en compte le plus grand espace entre les lignes verticales du spectrogramme des phrases prononcées en situation normale au cours du vol et ce même espace lors de la situation d'urgence. C'est un espace en unités métriques.

Pour chaque mot d'une phrase le VSSR est calculé, il en résulte un VSSR moyen pour la phrase avec son écart-type. Chaque phrase du vol est ainsi traitée. Les auteurs définissent neuf classes d'écart-type de 0 à 60% par tranche de 8,49% à l'exception des deux classes extrêmes qui ont une largeur de 0,25%. Les résultats sur l'ensemble des vols examinés montrent que 70% des voix en phase normale de vol appartiennent aux classes 1 à 3, 70% des voix en phase d'urgence sont comptabilisées dans les classes 4 à 6 et 70% de voix en phase d'urgence appartiennent aux classes 7 à 9. L'écart-type du VSSR semble donc apte à traduire l'évolution du stress émotionnel.

La mesure de l'espace entre les lignes verticales du spectrogramme est en fait une estimation de la fréquence fondamentale F_0 dont on ne retient ensuite que les valeurs maximales dans les mots prononcés. Les conclusions peuvent donc être interprétées en termes de F_0 et indiquent que ses fluctuations s'accroissent avec l'émotion du locuteur.

Dans une étude sur des communications radio entre des pilotes et des contrôleurs de la navigation aérienne au cours de situations où les circonstances donnaient une indication claire de l'émotion des locuteurs, il a été observé une augmentation systématique de la fréquence fondamentale médiane des distributions de valeurs (C.E.Williams, 1969). Pour quatre cas étudiés celle-ci augmente respectivement de 208 à 432 Hz, de 210 à 284 Hz, de 112 à 151 Hz, de 166 à 196 Hz. Le contour temporel de F_0 devient plus irrégulier et discontinu. On peut en déduire également que la dispersion des valeurs est accrue par l'accroissement de la largeur des distributions. L'analyse du récit en direct d'un journaliste assistant à l'embrasement du Zeppelin « Hindenburg » (cf page 7) montre également la hausse notable de F_0 et de sa gamme de variation (C.E.Williams, 1972).

La hausse de la fréquence fondamentale est confirmée par l'analyse de la voix d'un pilote militaire en situation d'urgence : 115 Hz en situation normale et 163 Hz en phase de stress (P.Benson, 1995). Elle est complétée par une hausse de la fréquence du premier formant des voyelles analysées (510 à 537 Hz). Les fréquences des deuxième et troisième formants restent stables mais la pente spectrale diminue de 2 dB par octave montrant un spectre plus « plat » en état de stress. Des mesures de distances spectrales sur la base du calcul de vecteurs cepstraux en échelle Mel (M.F.C.C) ne montrent pas de grandes différences entre les prononciations normales et celles en état de stress.

Un rapport d'accident aéronautique du bureau d'enquête américain des transports indique lui aussi une hausse graduelle de F_0 pour la voix du capitaine de 150 Hz à plus de 350 Hz (National Transportation Safety Board, 1994).

La perception du stress émotionnel a été étudiée pour la voix d'un pilote d'hélicoptère en situation normale et en situation de perte de contrôle de l'appareil (A. Protopapas, 1995 et 1997). Les mesures de jitter (APC, « Average Perturbation Contour ») n'ont pas permis d'estimer que le paramètre était sensible car les gammes de variation dans les deux états se superposaient totalement.

Ensuite, le principe de l'étude était de manipuler graduellement le jitter des signaux vocaux pour demander à des auditeurs de les noter de 1 (voix calme) à 7 (voix très stressée). Il s'avère que la voix normale a toujours été perçue comme telle de même que la voix stressée. La quantité de jitter ne provoque pas l'augmentation des scores dans les deux cas. Comme l'indiquent les auteurs, le jitter peut être un indicateur d'autres types de perturbations mais peut-être pas celles qui produisent un stress extrême et de grandes fluctuations de la fréquence fondamentale. Il est probablement adapté à distinguer des états de faible perturbation.

Pour s'assurer de la réelle perception des différences, il a été demandé aux auditeurs d'estimer le degré d'enrouement de la voix, toujours sur les mêmes échantillons de parole. Dans ce cas, les scores augmentent avec la quantité de jitter dans les deux situations témoignant de la perception des modifications du signal et de l'association de ce paramètre avec le caractère enroué de la voix.

Dans un deuxième type d'expérience, les auteurs ont étudié la perception de modifications de la fréquence fondamentale moyenne et de sa gamme de variation. Les signaux originaux montraient des F_0 plus élevées et des gammes plus larges ($F_{0\,max}$ - $F_{0\,min}$) en situation de stress. Une forte corrélation est mesurée entre la hausse artificielle des valeurs moyennes et maximales de F_0 et la perception du stress émotionnel. Ce n'est pas le cas pour la gamme de variation de la fréquence fondamentale lorsque ses modifications sont indépendantes des valeurs moyennes et maximales de F_0.

Enfin, l'écoute des enregistrements après inversion temporelle ne modifie pas les scores des auditeurs. Ce résultat suggère que le sens de variation de F_0 ne joue pas un rôle dans la perception du stress. Il s'agirait plutôt de la vitesse des fluctuations.

Pour estimer la fatigue de pilotes sur des vols long-courriers mais aussi celle de contrôleurs de la navigation aérienne, une publication montre deux voyelles [a] dans le plan des phases prononcées en situation normale et en état de fatigue (K.Shiomi, 2000). Les différences visuelles sont quantifiées par la mesure du premier exposant de Lyapunov. Les auteurs suggèrent qu'il s'agit d'une piste pour la détection. Ils ont développé leurs travaux hors du cadre des transports (§ 2.3 page 41).

Une dernière étude effectue la comparaison entre la voix d'un pilote issue de l'enregistreur phonique et celle d'une simulation en laboratoire. Elle conclu que la cause de l'accident est due à l'hypoxie c'est-à-dire à la diminution de la quantité d'oxygène distribuée aux tissus, par exemple lors d'une montée trop rapide en altitude (I.Saito, 1980). Les caractéristiques observées sont le temps d'établissement du voisement (« Voice Onset Time ») et la fréquence fondamentale.

▪ ***Simulation de vol :*** Le simulateur de vol est un outil qui fournit matière à étudier divers types de perturbations. En témoigne par exemple l'effet positif d'une tâche complexe de pilotage sur diverses grandeurs physiologiques (rythme cardiaque, pression sanguine, activité respiratoire, clignement des yeux) (J.A.Veltman, 1996).

Il peut également servir à simuler de longues charges de travail (J.Whitmore, 1996). Dans cette étude, trois missions de 36 heures chacune avec 36 heures de repos entre elles ont été analysées vocalement puis complétées d'une auto-évaluation subjective de la fatigue par chaque pilote et de la réalisation de tests cognitifs.

Les résultats montrent un effet significatif de la mission sur la durée moyenne des mots d'une même phrase prononcée toutes les trois heures par deux pilotes. Elle diminue de 453ms à 426ms entre les deux premières missions et vaut 413ms pour la dernière.

La fréquence fondamentale est moyennée pour la prononciation d'un des mots de la phrase. Elle est de 114,8 Hz pour la mission 1, puis de 111,2 Hz pour la seconde et de 113,9 Hz pour la troisième. Les différences sont faibles et non significatives. En revanche, au cours d'une même mission, les variations observées sont beaucoup plus importantes (de 108 Hz à

118 Hz par exemple) et aboutissent à des modifications significatives d'une prononciation à l'autre. C'est également le cas pour la durée moyenne des mots.

Les résultats des tests cognitifs et de l'auto-évaluation de la fatigue corroborent les modifications vocales mesurées. Par exemple, l'accroissement de fatigue perçu en fin de mission s'accompagne par la baisse sensible de la fréquence fondamentale.

Les auteurs concluent que la parole peut être un indicateur de l'état de fatigue du locuteur.

Dans une étude similaire réalisée en simulateur avec trois pilotes, la tâche a consisté à poser quatre fois un appareil dans trois conditions de vent et de turbulences différentes fournissant ainsi divers degrés de charge de travail (durée 15 minutes) (J.Schneider, 1989). Les analyses vocales des communications avec le contrôle aérien étaient complétées de scores subjectifs où le pilote estimait toute les 90 secondes par une note de 1 à 10 le niveau de difficulté de la procédure.

Trois mesures acoustiques ont été effectuées : l'amplitude relative moyenne, la fréquence, la durée du pic d'amplitude d'une syllabe. Aucune information supplémentaire n'est fournie sur le vocabulaire dont sont extraites ces mesures ni sur la méthode d'analyse. On peut supposer qu'il s'agit de la fréquence fondamentale mais on ne sait pas de quelle amplitude il s'agit, ni à quelle syllabe fait référence la mesure de durée et comment ces grandeurs sont mesurées.

Les conclusions indiquent que pour deux pilotes sur trois la hausse de charge de travail se traduit par une plus grande amplitude et une plus grande durée des pics malgré la grande variabilité des résultats.

Le simulateur a été utilisé dans un autre contexte militaire de réalisation de missions de détection, d'identification, d'interception et de destruction d'appareils (J.J.Congleton, 1997). L'objectif était d'étudier la fréquence fondamentale, le jitter et le shimmer de la voix de seize pilotes dans différentes conditions de charge de travail estimées par un modèle spécifique. Ce dernier définit cinq niveaux de charge sur la base des scénarios simulés. Les résultats montrent un accroissement significatif de la fréquence fondamentale (de 130, 145, 150, 153 à 155 Hz) et une décroissance du jitter (de 1.25, 1.2, 0.95, 0.8 à 0.9%) avec l'augmentation du niveau de charge. Les mesures de shimmer ne montrent pas de tendance particulière avec la hausse du stress provoqué.

Dans un contexte similaire, une analyse linguistique du discours de pilotes de chasse a été élaborée pour isoler et identifier les énoncés porteurs d'émotion : neutre, situation de stress, complicité, interjections et exclamations, hésitations (F.Loyau, 2007).

Une dernière étude utilise un appareillage qui s'apparente à un simulateur fixe servant à maintenir deux curseurs à l'écran à l'aide d'un manche (premier curseur) et de pédales (second curseur) (G.R.Griffin, 1987).

Un deuxième exercice consiste à diffuser au casque des séquences de nombres et de chiffres différentes à chaque oreille. Le pilote doit reporter ou annoncer la séquence entendue dans l'oreille qui lui a été désignée.

Le troisième niveau de difficulté consiste à cumuler l'exercice psychomoteur et la tâche d'écoute dichotomique. 60 sujets ont été testés pour plus de 50 essais dans chaque configuration.

Les réponses orales des locuteurs ont été analysées pour mesurer la fréquence fondamentale au milieu de la prononciation, son niveau sonore maximal et la durée moyenne des mots.

Il en résulte que la fréquence fondamentale croit avec la complexité des tâches et pour chacune significativement par rapport à la première. Les variations sont faibles (de 106,69 à 118,91 Hz) mais rendues significatives par le nombre de mesures (plus de 1200 par sujet). C'est également le cas pour le niveau sonore maximal qui augmente lui aussi, et pour la durée moyenne des mots qui diminue. Dans la mesure où des estimations presque simultanées du niveau sonore et de la fréquence fondamentale ont été réalisées, on peut regretter l'absence d'étude de la corrélation entre les deux paramètres.

▪ ***Base de données SUSAS :*** ("Speech Under Simulated and Actual Stress"), Elle est constituée d'une majorité d'enregistrements sonores s'inscrivant dans le contexte aéronautique. Ils sont composés d'un vocabulaire issu du secteur et simulent des situations comme l'effet Lombard, l'effet de charges de travail psychomotrices, les effets de l'accélération mais aussi la peur, l'anxiété, la colère, des styles de parole (lente, rapide, forte …) et enfin des conversations à bord d'un hélicoptère (J.H.L.Hansen 1996 et 2000). L'auteur qualifie toutes ces prononciations de « stressées ».

Les travaux sur la base de données ont eu pour objectif de proposer et d'améliorer les performances d'un système de classification et de reconnaissance automatique de ces « stress ». Les descripteurs acoustiques utilisés sont articulatoires (rapport des aires du conduit vocal), prosodiques (fréquence fondamentale, intensité), euclidiens (distance entre ensembles de coefficients cepstraux et entre valeurs moyennes des autres paramètres) et spectraux (formants). Les taux de reconnaissance obtenus sont très variables suivant la situation considérée (J.H.L.Hansen 1994 et 1996, B.D.Womack 1996 et 1999).

Un paramètre acoustique original a été introduit parmi l'ensemble des descripteurs utilisés pour la reconnaissance : le « Teager Energy Operator » (TEO) (D.A.Cairns, 1994). Il s'apparente à une mesure de l'énergie échantillon par échantillon. Il opère sur le signal de parole filtré passe bande dans la zone de fréquence d'un formant pour un segment temporel compris entre deux impulsions glottiques. L'opérateur témoignerait de l'existence de composantes non linéaires dues à des tourbillons aérodynamiques dans la zone des fausses cordes vocales. Les auteurs montrent que le TEO subi des modifications en situation de stress. L'évolution temporelle du TEO aboutit à un profil distinct suivant les situations de parole « stressée » étudiées (D.A.Cairns, 1994).

Il a été testé dans un cas d'incident aéronautique réel comme descripteur acoustique d'un classificateur automatique sans aboutir à une reconnaissance effective du degré de stress (G.Zhou, 1999).

Plus récemment, des mesures dérivées du TEO ont été proposées. Elles effectuent le calcul de l'aire sous l'enveloppe de la fonction d'autocorrélation des TEO obtenus en sortie d'un banc de filtre en bandes critiques (G.Zhou, 2001). Ce descripteur apporte des améliorations à la reconnaissance automatique des situations de la base de données SUSAS et a été appliqué pour détecter le stress de soldats en situation d'évaluation orale devant un jury (M.A.Rahurkar, 2002 et 2003).

Transport Maritime

Les deux études en la matière traitent de l'accident du pétrolier « Exxon Valdez » le 24 Mars 1989 semble-t-il dû à l'état d'alcoolémie du capitaine. Les auteurs ont travaillé sur les enregistrements des conversations entre ce dernier et les gardes côtes. Les résultats des analyses suggèrent que le capitaine était en état d'ébriété au moment de l'accident.

Les paramètres mesurés sont la durée et la fréquence fondamentale moyenne de segments de parole. Le locuteur s'exprime plus lentement avec une fréquence fondamentale plus basse et avec une augmentation de sa variabilité accompagnées de défauts d'articulation (M.Brenner 1991, J.A.Tanford, 1991).

Transport Ferroviaire

Les travaux des auteurs ont pour objectif d'estimer la vigilance de conducteurs de train par analyse vocale (P.Shahidi, 2008, 2009 et 2010). C'est un travail sur la détection de la fatigue. Les indicateurs acoustiques sont exclusivement temporels. Il s'agit des durées minimales, maximales, moyennes des mots et des pauses par phrase, des pauses entre les phrases, du débit (moyenne de la durée des mots pour la prononciation d'un nombre fixe de mots). Ces mesures sont accompagnées de celle du niveau sonore des mots et du rapport signal sur bruit.

Les valeurs numériques des paramètres sont traitées par un algorithme d'intelligence artificielle (logique floue) où figurent des règles établies empiriquement pour statuer sur le degré de fatigue (très faible, faible, normal, élevé, très élevé) à partir des valeurs numériques des descripteurs acoustiques précédents mesurés en quasi temps réel. Le système fournit alors un quotient indicatif du niveau de fatigue du conducteur de train.

En l'absence de plus d'information sur le traitement effectué, et surtout d'expériences de validation dans des conditions de laboratoire accompagnées d'un contrôle et de mesures psychophysiologiques de la fatigue sur une population suffisante, il est difficile d'évaluer la fiabilité du système automatique d'analyse. C'est néanmoins une étude qui statue sur une perturbation de l'état général du locuteur détectable sur la seule base de mesures acoustiques autour du débit phonatoire.

D'autres travaux statuent aussi sur la capacité de détection de la fatigue de conducteurs de trains mais fondés cette fois-ci sur l'estimation du caractère chaotique de la voix (K.Shiomi, 2008 et 2009, § 5.2.3 page 139). L'examen de ces études est présenté au paragraphe 2.3 car les mesures traitent de la fatigue au-delà du secteur des transports.

Transport Automobile

Des conducteurs automobiles ont pour tâche d'effectuer des opérations arithmétiques pendant leur conduite sur un simulateur, ceci à deux vitesses données et avec deux contraintes de temps pour s'exprimer (R.Fernandez, 2003). L'objectif est de mettre en place un système de reconnaissance automatique qui discrimine par analyse vocale les quatre combinaisons expérimentales.

Elles définissent quatre états de stress ou plus exactement quatre charges cognitives provoquées par les tâches mentales.

Les performances de plusieurs méthodes de classification sont testées et comparées pour la reconnaissance automatique de ces états. Elles aboutissent à des scores moyens légèrement supérieurs à 50% sur la base d'un calcul de l'opérateur non linéaire de Teager pour 21 bandes de fréquence du spectre 0 – 4 kHz.

On peut rajouter à cette étude le résumé de travaux sur la détection audiovisuelle d'émotions pour des conducteurs automobiles (B.Schuller, 2008). Sont utilisés les paramètres acoustiques suivants : fréquence fondamentale, énergie, enveloppe, fréquence amplitude et

largeur de bande des premiers formants, seize coefficients M.F.C.C, schimmer, jitter. Mais rien n'est indiqué sur leur poids, leur sensibilité et l'apport des données audio à la reconnaissance visuelle de l'émotion sur les visages.

Enfin, l'interaction de tâches cognitives avec la conduite automobile en parcours urbain (interactions avec un passager et avec un système automatique de dialogue pour deux situations émotionelles) a montré que des paramètres acoustiques variaient. La fréquence fondamentale moyenne, le centre de gravité spectral et la durée moyenne des segments voisés augmente entre le dialogue avec le passager et celui avec le système automatique. Les mêmes tendances apparaissent entre le discours neutre et le discours émotionnel avec la machine (hésitations, confusions, colère). Ces résultats, associés à des coefficients cepstraux ont été appliqués pour servir de base à une reconnaissance automatique de la charge cognitive et ont conduit à obtenir un taux de plus de 90% (H.Bořil, 2010).

Les travaux réalisés dans le domaine des transports sont en petit nombre malgré le recensement sur une grande période de temps. Ils témoignent tout de même de l'intérêt des organismes concernés.

Des informations importantes manquent souvent dans les publications pour avérer les résultats obtenus. Il n'apparait pas de méthode et d'indicateurs qui puisse statuer sur l'état de perturbation du locuteur. En revanche, sa détection semble pouvoir être effectuée par certains paramètres. Ils résident essentiellement autour de la fréquence fondamentale (valeurs moyennes, maximales, dispersion …).

Peu d'études traitent un nombre important de sujets certainement à cause de l'absence de partenaires du secteur qui s'engagent sur des expérimentations à plus grande échelle. Les conditions sont telles qu'elles nécessitent une forte implication des partenaires publics ou privés qui possèdent les appareillages nécessaires pour les expérimentations et peuvent intervenir pour trouver les sujets qui conviennent.

Cette recherche dans le secteur des transports est assez peu développée.

2.2. Voix et Émotion (références bibliographiques § 5.2.2. page 126)

L'appellation « émotion » ne figure pas toujours en tant que telle dans les travaux examinés. La littérature anglo-saxonne utilise diverses expressions comme « emotional stress », « emotional speech » mais aussi « psychological stress », « affect », « stress ».

Pour le psychologue spécialiste des émotions tous ces vocables peuvent désigner des phénomènes différents conduisant à catégoriser les expérimentations correspondantes. Mais il n'apparaît pas de distinction conceptuelle claire dans les travaux des acousticiens.

Parfois, sous l'appellation « stress » il est question aussi bien d'expression vocale dans un environnement bruyant (effet Lombard), que de mensonge, d'expression d'émotions primaires, de situations de danger ou d'évènements négatifs. Il y a néanmoins consensus pour indiquer que le stress comprend une série de modifications physiologiques et comportementales qui préparent l'organisme pour une réponse appropriée (K.R.Scherer, 1981). Seule l'expression « cognitive stress » paraît être réservée à des expérimentations du type du test de Stroop ou à la réalisation de tâches arithmétiques qui mettent en œuvre des inférences cognitives.

La définition et la description des états émotionnels constituent une problématique à part entière pour le psychologue (R.Cowie 2000 et 2003, R.R.Cornelius, 2000 ; K.R.Scherer, 2003).

Pour le non spécialiste en Psychologie, tous les travaux relatifs à ces expressions sont regroupés ici sous le terme générique d'« émotion ».

Le raisonnement est simple et pragmatique. Il consiste à considérer que lorsque le ou les facteurs perturbateurs du locuteur touchent à la sphère affective, les expérimentations s'inscrivent dans ce paragraphe « voix et émotion » même si elles induisent également des modifications physiologiques.

Lorsque le facteur est clairement identifié comme de la fatigue d'origine physique ou « nerveuse », les travaux figurent dans le paragraphe « voix et fatigue ».

Viennent ensuite des manifestations vocales dont le contexte permet d'éviter au non spécialiste de les classer : le mensonge, les pathologies non vocales et l'état d'alcoolémie.

En définitive, on opte pour un classement par type d'expérimentation permettant ainsi au psychologue de les rattacher à la typologie qui lui est familière.

Ainsi, pour le paragraphe intitulé « voix et émotion », sont présentés les travaux relatifs à la simulation d'émotions primaires, à l'analyse de situations de travail réelles au fort contexte émotionnel, au test de Stroop, à la réalisation de tâches arithmétiques.

Il s'agira d'analyse du signal vocal et lorsque c'est le cas, de synthèse sonore et de reconnaissance automatique de la parole.

L'effet Lombard n'est pas traité dans cette revue bibliographique car il n'est pas générateur d'émotion. En revanche, l'activité professionnelle en milieu bruyant peut contribuer à la fatigue générale du locuteur. Il ne s'agit pas là de l'effet lui-même mais des conditions de travail. L'effet du bruit, lorsqu'il existe, intervient alors au même titre que la pénibilité ou la durée, comme facteur contributif de la fatigue.

C'est Charles Darwin en 1872 qui cite et complète de son analyse et de ses observations les travaux de Herbert Spencer en 1858 : « le caractère que prend la voix humaine sous l'influence de diverses émotions a été examiné par M. Herbert Spencer dans son intéressant essai sur la musique. Il montre clairement que la voix se modifie beaucoup suivant les circonstances, dans sa force et dans sa qualité, c'est-à-dire dans sa sonorité et dans son timbre, dans sa hauteur et dans son étendue. Personne ne peut écouter un orateur ou un prédicateur éloquent, ou un homme s'adressant avec colère à un autre, ou quelqu'un exprimant de l'étonnement, sans être frappé de la justesse des remarques de M. Spencer. Il est curieux de constater que l'inflexion de la voix devient expressive très tôt dans la vie ».

Il est un des premiers à mettre en relation la fréquence fondamentale et les émotions : « Il est assez clair que la hauteur de la voix est en relation avec certains états de la sensibilité. Une personne qui se plaint doucement de mauvais traitement, ou qui souffre légèrement, parle presque toujours d'une voix que l'on dit haut perchée » (C.Darwin, 2001 dans §5.2.5 page 142).

« Avant lui le musicologue Friedrich Marpurg avait essayé d'établir un lien entre schémas acoustiques spécifiques et états émotionnels particuliers. Le chagrin par exemple, se caractérisait par une mélodie lente et dissonante et l'envie, suggérait-il par des tons grondeurs et irritants » (A.Karpf, 2008 dans §5.2.5 page 142).

Depuis, les moyens d'analyse se sont considérablement développés jusqu'à proposer aujourd'hui des systèmes de reconnaissance automatique de l'émotion. De nos jours, les multiples situations qui peuvent provoquer des désordres vocaux sont identifiées. Le scientifique peut en conserver la trace sonore par l'enregistrement et il dispose de banques de données qui les répertorient dans différentes langues. On peut citer par exemple la « Berlin Emotional Database » pour l'allemand, Speech Under Simulated and Actual Stress (SUSAS) et Belfast Database pour l'anglais, « Danish Emotional Speech Database » pour le danois (E.Douglas-Cowie, 2003). Elles servent à rechercher les paramètres acoustiques sensibles à l'émotion et à tester les appareillages de reconnaissance automatique. La plupart d'entre elles utilisent des acteurs pour simuler l'émotion mais certaines possèdent de la parole prononcée lors de situations réelles (N.Campbell, 2000).

Parfois des tests d'écoute sont mis en place pour estimer le caractère naturel des émotions enregistrées ou synthétisées. Les aspects perceptifs de l'émotion que Charles Darwin a décrit pour l'homme et les animaux sont devenus un point d'appui pour la confirmation des résultats des analyses, de la reconnaissance et de la synthèse en complément de mesures physiologiques. C'est aussi le cas pour les aspects purement visuels comme les modifications du visage sous l'effet de l'expression des émotions et les différentes recherches audiovisuelles qui en découlent (R.Cowie 2001, B.Granström 2005, T.Bänziger 2009, M.Wöllmer 2010).

Ce qui au 19ème siècle relevait de l'observation et de l'écoute est aujourd'hui objectivement démontré. L'expression vocale et les paramètres acoustiques qui la caractérisent sont porteurs d'une information sur l'émotion du locuteur (J.A.Bachorovski, 2008). Les observations de C.Darwin sur les animaux sont également étudiées (C.Fichtel 2001, J.Soltis 2009). Dès les années 30 des auteurs se sont penchés sur les indicateurs vocaux de l'émotion permettant de repérer des signes de variation essentiellement de la fréquence fondamentale, du niveau sonore et du rythme (K.R.Scherer, 1981).

La littérature est devenue très abondante aussi la présentation est plus synthétique que pour le secteur des transports et s'attache à rendre compte des paramètres acoustiques mesurés et des résultats généraux obtenus.

2.2.1. Émotions Simulées ou Induites

Les psychologues distinguent six émotions primaires présentes dès la naissance, universelles et déterminées génétiquement : la peur, la colère, le dégoût, la tristesse, la joie, la surprise. Les émotions secondaires sont beaucoup plus nombreuses, apprises et influencées par la culture (honte, fierté, jalousie …).

Les travaux recensés dans ce paragraphe traitent de l'expression vocale lors de la manifestation essentiellement des émotions primaires lorsqu'il est demandé à des locuteurs de les simuler (acteurs professionnels ou non) ou que le protocole expérimental les provoque. Il semble d'ailleurs que les émotions jouées soient plus fortement perçues par des auditeurs que les émotions induites, suggérant le fait que la simulation peut conduire à sur jouer la situation émotionnelle (J.Wilting, 2006).

Le principe des études consiste à analyser le signal vocal pour rechercher des corrélations acoustiques ou à effectuer une reconnaissance automatique de l'émotion par des procédures algorithmiques spécifiques. Il peut également s'agir de réaliser une parole synthétique au contenu émotionnel.

2.2.1.1. Émotions Simulées par des Acteurs

Deux approches existent : celle de l'analyse du signal pour rechercher des paramètres acoustiques susceptibles de devenir des indicateurs de l'émotion, et celle de la reconnaissance automatique, où les paramètres sont plutôt des descripteurs du signal que les méthodes et les algorithmes de classification utilisent pour parvenir à un score de reconnaissance.

2.2.1.1.1. Analyse Acoustique

Divers paramètres ont été mesurés en analysant le signal vocal d'acteurs simulant des émotions. La synthèse bibliographique montre que certaines caractéristiques possèdent en moyenne des valeurs numériques différentes selon l'émotion considérée. Cette conclusion générale qui stipule que le signal vocal est modifié par l'émotion jouée est d'ailleurs interculturelle puisqu'elle est vérifiée quelle que soit les langues étudiées :

Allemand (R.Banse 1996, M.Kienast 2000, K.Alter 2000 et 2003, A.Paeschke 2000 et 2004, T.Bänziger 2005, K.Hammerschmidt 2006, N.Kamarudin 2009) ;
Anglais (G.Fairbanks 1938, P.Lieberman 1961 et 1962, C.E.Williams 1972, L.Cosmides 1983, R.Cowie 1996, S.P.Whiteside 1998, C.Pereira 1998, V.A.Petrushin 2004, S.Lee 2005, H.S.Cheang 2008) ;
Chinois (J.Yuan 2002) ;
Coréen (T.Li 2005) ;
Français (M.M.Homoyampour 1994, K.R.Scherer 1998, G.Cédric 2002, T.Ehrette 2002, G.Beller 2008) ;
Japonais (N.D.Cook 2003, T.Fujisawa 2004, C.Menezes 2006, H.Kawatsu 2007, S.Takeda 2008) ;
Malais (A.A.Razak 2003) ;
Néerlandais (S.J.L.Mozziconnacci 2000) ;
Russe (P.V.Simonov 1975) ;
Tchèque (A.Novak 1993) ;
mais aussi lors de prononciations de mots sans sens (A.M.Laukkanen 1996, M.Goodbeek 2008 et 2009).

Les descripteurs acoustiques qui présentent des modifications notables sont les suivants. Ils n'ont pas tous été mesurés dans toutes les langues précédentes.

Paramètres Prosodiques

La fréquence fondamentale F_0 est le premier d'entre eux. On en mesure la valeur moyenne (ou plus rarement la valeur médiane) des structures prononcées (voyelles, syllabes, mots, phrases) mais aussi le contour temporel au cours de la prononciation et la dispersion (différence entre valeurs maximale et minimale, écart-type des distributions, jitter). C'est le paramètre le plus universellement mesuré.

Tous les auteurs s'attachent à démontrer que les émotions simulées entrainent des variations de F_0 par rapport à l'état de neutralité émotionnelle et proposent parfois de hiérarchiser les émotions selon leur fréquence fondamentale moyenne. Il apparaît ainsi que F_0 augmente en général depuis l'état neutre et ceux de l'expression de la tristesse ou de l'ennui

jusqu'à l'expression de la colère ou de la peur. À l'appui des références qui suivent, la fréquence fondamentale moyenne possède des valeurs numériques distinctes d'une émotion à l'autre et différentes de la neutralité (G.Fairbanks 1938, R.Bonner 1943, C.E.Williams 1972, L.Cosmides 1983, M.Homoyampour 1994, R.Banse 1996, A.M.Laukkanen 1996, C.Pereira 1998, K.R.Scherer 1998, S.P.Whiteside 1998, J.Yuan 2002, A.A.Razak 2003, S.Lee 2005, T.Li 2005, C.Menezes 2006, H.S.Cheang 2008).

La dispersion des valeurs de F_0 est très majoritairement modifiée lors de l'expression vocale des émotions simulées (mêmes références que les précédentes et P.Lieberman 1961, J.Yuan 2005, T.Bänziger 2005).

Grace aux nombreuses approches mises en œuvre pour quantifier les variations observées du contour temporel de la fréquence fondamentale au cours d'une prononciation, la synthèse des analyses peut conclure à l'existence de modifications des contours en situation émotionnelle. On peut également noter que lorsque plusieurs niveaux d'émotion sont simulés, les contours se décalent vers des valeurs de F_0 plus grandes et se déforment d'autant plus que ce degré émotionnel est important (K.Hirose, 2000). Il n'est pas pour autant établi de contour type pour chaque émotion (C.E.Williams 1972, L.Cosmides 1983, C.Pereira 1998, A.Paeschke 2000 et 2004, T.Bänziger 2005, H.Kawatsu 2007, C.T.Ishi 2006 et 2008).

Enfin, l'étude de la forme de l'onde glottique a montré des différences quantitatives significatives selon les émotions (K.E.Cummings 1995, A.M.Laukkanen 1996, G.Cédric 2002, S.R.M.Prasanna 2010), prouvant ainsi qu'elles ont un effet sur la source d'excitation de l'appareil phonatoire.

Paramètres Articulatoires

Mis en évidence à partir de l'analyse spectrale, ils concernent principalement la fréquence des deux premiers formants. Les variations mesurées ne sont pas toujours significatives comme pour la plupart des paramètres prosodiques. On peut néanmoins conclure à des modifications sous l'effet de la simulation d'émotions (M.Homoyampour 1994, M.Kienast 2000, D.Erickson 2000, S.Lee 2005, G.Beller 2008, M.Goodbeek 2009, G.Salvi 2010).

Des différences articulatoires ont également été mises en évidence entre les différentes émotions en étudiant la vitesse de la pointe de la langue en fonction de sa position (S.Lee 2005, D.Erickson 2000, J.Kim 2009).

Les analyses spectrales montrent des modifications de la répartition de l'énergie grâce :

- à l'estimation de rapports énergétiques ou de différences de niveau sonore entre différentes bandes de fréquence (R.Banse 1996, R.Cowie 1996, K.R.Scherer 1998, H.Cheang 2008) ;
- aux mesures de pente spectrale (K.Alter 2000, J.Yuan 2002, A.Novak 1993), à celles de centroïdes (M.Kienast 2000, C.Menezes 2006, A.Novak 1993) ;
- à celles résultant de la simple observation de sonagrammes (C.E.Williams 1972, T.Suzuki 2008) ;
- et à la mesure de paramètres spécifiques comme la fréquence possédant l'amplitude la plus élevée dans le spectrogramme (K.Hammerschmidt 2007).

Enfin, sur la base de coefficients spectraux, des distances métriques ont pu être établies entre différentes émotions permettant de visualiser leur position sur un diagramme à deux dimensions (K.P.Truong, 2007). Des mesures de dissimilarité entre les enveloppes

spectrales ont montré qu'elles se modifiaient pour l'expression de la tristesse et de la joie par rapport à celles de l'état neutre (G.Salvi, 2010).

Paramètres de Niveau Sonore et Temporels

L'amplitude du signal a fait l'objet d'études qui démontrent son lien avec la simulation des émotions (L.Cosmides 1983, M.Homoyampour 1994, R.Banse 1996, R.Cowie 1996, C.Pereira 1998, K.R.Scherer 1998, A.A.Razak 2003, C.Menezes 2006, K.Hammerschmidt 2007, H.S.Cheang 2008). Il s'agit également du schimmer (S.P.Whiteside 998, T.Li 2005). Les sens de variation observés sont le plus souvent naturellement corrélés à ceux de la fréquence fondamentale.

Dans le domaine temporel, plusieurs types de mesures ont été effectués comme celle de la durée des prononciations, celle des pauses ou celles du débit phonatoire. Elles traduisent des raccourcissements ou des allongements de l'expression selon l'émotion simulée : par exemple une expression plus lente pour la tristesse que pour la colère (R.Bonner 1943, C.E.Williams 1972, L.Cosmides 1983, M.Homoyampour 1994, R.Banse 1996, R.Cowie 1996, C.Pereira 1998, K.R.Scherer 1998, K.Hirose 2000, S.J.L.Mozziconacci 2000, J.Yuan 2002, A.A.Razak 2003, K.Hammerschmidt 2007, H.Kawatsu 2007, H.S.Cheang 2008).

Une étude interculturelle entre les mêmes émotions simulées pour l'allemand et le coréen montre que les variations de la durée de segments phonétiques dépendent non seulement de la langue mais aussi de l'émotion jouée (M.Goudbeek, 2010).

Le rapport H.N.R (« Harmonic-to-Noise Ratio »), qui exprime la proportion du niveau des harmoniques dans le signal par rapport au niveau de bruit qu'il contient, est généralement associé à l'estimation de la qualité de la voix. Il a été évalué pour l'expression d'émotions simulées (K.Alter 2000 et 2003, T.Li 2005, H.S.Cheang 2008, S.Takeda 2008). Les auteurs observent des variations significatives du rapport entre les différents états émotionnels. Elles peuvent conduire à les ordonner (S.Takeda 2008) et à conclure à de faibles variations (K.Alter 2000 et 2003, T.Li 2005, H.S.Cheang 2008).

En définitive, la littérature ne propose pas de valeurs numériques seuils, éventuellement associées à des intervalles de confiance, ou encore de pourcentages moyens de variation par rapport à l'état neutre afin de différencier les effets des émotions vocalisées.

Mais la juxtaposition des variations de nombreux paramètres précédents a permis d'établir des profils acoustiques individuels des émotions sur la base d'analyses discriminantes, d'analyse de variance et de corrélation entre les descripteurs (R.Banse 1996, K.R.Scherer 1998, K.Hammerschmidt 2007) ou sur la simple base des résultats conjoints des mesures (G.Fairbanks 1938). La critique qui peut en être faite est que, si le traitement statistique permet d'établir une classification, l'attribution du profil d'une nouvelle vocalisation à une catégorie d'émotion n'est pas démontrée. Les profils restent en effet très voisins avec très certainement des intervalles de confiance qui se chevauchent. Sans un traitement statistique approprié la classification est difficile à la simple lecture des valeurs des paramètres. Les résultats fournissent donc un ensemble de descripteurs acoustiques des états émotionnels.

Par ailleurs bien qu'aucun paramètre acoustique ne porte à lui seul l'information affective même s'il est rattaché à ses valeurs statistiques, la fréquence fondamentale possède un poids important puisqu'elle est quasi systématiquement sujette à des modifications significatives lors d'une expression vocale émotionnelle.

2.2.1.1.2. Reconnaissance Automatique

La finalité des recherches est de réaliser un outil informatique dont le taux de reconnaissance de l'émotion soit le plus élevé possible. Les systèmes en question utilisent des descripteurs acoustiques regroupés en vecteurs pour procéder au traitement (les principaux sont décrits dans l'article de D.Ververidis, 2006).

Les systèmes opèrent une classification automatique qualifiée de supervisée lorsque les classes sont fournies au programme informatique avec les données acoustiques, ou de non supervisée lorsqu'elles sont déterminées automatiquement à partir des données fournies.

L'apprentissage, qui est la procédure permettant de caractériser les propriétés acoustiques de chaque classe d'émotion, peut être dépendant ou indépendant du locuteur. Il est dépendant lorsque le classificateur est entrainé et exécuté sur des échantillons de parole qui proviennent du même locuteur et il est indépendant lorsqu'ils proviennent de différents locuteurs. Les performances chutent en l'absence de référence au locuteur considéré.

Généralement la reconnaissance de l'émotion est non supervisée et un corpus d'apprentissage ou d'entrainement est fournit aux algorithmes pour qu'ils génèrent la classification. Puis un second corpus de parole, différent du premier, permet de mesurer la performance de la reconnaissance. Elle s'exprime comme le pourcentage d'éléments de parole prononcés dans une situation émotionnelle simulée et reconnus comme tels par le système.

Les techniques de classification sont fondées sur des modèles de Markov cachés (HMM ou « Hidden Markov Models »), des réseaux de neurones artificiels (ANN ou « Artificial Neural Networks »), des analyses discriminantes sur les densités de probabilité des descripteurs (GMM ou « Gaussian Mixture Model »), des mesures de distance (KNN ou « K-Nearest-Neighbors) ou encore sur la séparation des vecteurs par des frontières (SVM ou « Support Vector Machines ») (D.Ververidis 2006).

L'objectif des travaux est double :

- améliorer les performances en introduisant de nouveaux paramètres acoustiques ou en recherchant l'ensemble de descripteurs qui conduit aux meilleurs résultats ;
- améliorer les performances des algorithmes et des méthodes de reconnaissance.

Il y a donc complémentarité entre l'approche analyse et l'approche reconnaissance des émotions. En revanche, les conditions de la reconnaissance des émotions diffèrent de celles de la reconnaissance du locuteur. Pour cette dernière, les ensembles de descripteurs doivent présenter une forte variabilité inter - locuteur et une faible variabilité intra – locuteur. Celle-ci est d'ailleurs compromise par ses états psychologiques, physiologiques et les styles de parole. Avec la reconnaissance des émotions, la variabilité intra doit être forte pour en favoriser la distinction et la variabilité inter faible.

Depuis une dizaine d'années, l'approche reconnaissance automatique des émotions a pris le meilleur sur les analyses acoustiques. Une des raisons possibles est l'échec de l'analyse à fournir des paramètres sensibles et fiables avec des valeurs seuils et des gammes de variation spécifiques. La recherche a établit des constats et elle a démontré que le signal variait significativement lorsqu'une émotion était simulée. L'étape de la modélisation n'a pas

été franchie. Parallèlement, l'essor des techniques de reconnaissance automatique du locuteur a été considérable et son application à la reconnaissance des émotions en a fortement bénéficié. Par ailleurs, la prise en compte de l'émotion dans les systèmes opérant une vérification du locuteur s'avère nécessaire car elle provoque une diminution de la performance des systèmes par l'accroissement de la variabilité intra-locuteur (W.Wu, 2006).

Même si le cadre de cette étude concerne l'analyse du signal vocal, les résultats en matière de reconnaissance ne sauraient être ignorés. Ils sont décrits ici pour la simulation d'émotions comme ils le seront, dans les prochains paragraphes, pour les émotions induites par le contexte expérimental et les cas réels.

L'objet n'est pas d'entrer dans le détail des méthodes de classification mais de livrer une synthèse des performances sur la base des nombreux descripteurs choisis par les auteurs complétant de fait, les éléments d'analyse acoustique précédents.

Les vecteurs acoustiques sont constitués de ces descripteurs généralement accompagnés de leurs valeurs statistiques (moyenne, minimum, maximum, écart-type, paramètres des distributions…). La reconnaissance est sensiblement améliorée par la présence des données statistiques (V.Hozjan, 2003). Quelles que soient les langues, la majorité des systèmes de reconnaissance utilisent des paramètres de la prosodie, énergétiques et relatifs à la durée des prononciations. Ils leurs associent parfois d'autres caractéristiques, par exemple spectrales. Les références bibliographiques relatives aux descripteurs les plus classiques comme la fréquence fondamentale ne sont pas exhaustives : ce sont des exemples de travaux où les paramètres en question sont mesurés. On les retrouve dans un grand nombre d'autres études.

Comme il sera vu par la suite, les travaux sur la synthèse sonore s'accompagnent parfois d'une analyse acoustique des émotions afin d'en rechercher les paramètres pertinents pour la coloration émotionnelle d'un discours neutre. À titre d'exemple, une étude propose pour chaque émotion simulée des gammes et des sens de variation pour plusieurs paramètres acoustiques essentiellement prosodiques (I.Iriondo, 2000). Un second travail de recherche introduit des pourcentages de variation de plusieurs caractéristiques (F_0 moyenne, gamme de variation de F_0, durée, intensité, jitter, shimmer, paramètres de la source glottique) à appliquer pour générer chaque émotion primaire dans la parole (J.P.Cabral, 2006).

Descripteurs Prosodiques

La fréquence fondamentale et son contour temporel sont systématiquement utilisés par les classificateurs. La grande majorité des publications référencées en la matière traitent la reconnaissance des émotions en prenant en compte la fréquence fondamentale. De fait, la liste est longue et n'est pas mentionnée en tant que telle mais, à de rares exceptions près, toutes les publications citées par ailleurs intègrent la fréquence fondamentale comme descripteur.

Des variantes existent pour quantifier le contour : dérivée première ou pente, gamme de variation, valeurs maximales, vitesse d'augmentation ou de décroissance (T.Seppänen, 2003) estimation de coefficients de régression (R.Tato, 2002), comptage des inflexions (D.Ververidis, 2005), statistiques sur le lissage et la dérivée première des contours (F.Dellaert, 1996). Le jitter intervient également (S.Yacoub, 2003). Des paramètres de la forme d'onde glottique interviennent aussi comme descripteurs acoustiques (quotients d'ouverture, de fermeture et de vitesse) (A.I.Iliev 2011, E.Väyrynen 2011).

Descripteurs Énergétiques

Rares sont les études qui précisent le calcul et l'unité des paramètres. Le terme énergie est donc maintenu ici faute de précisions sur les durées d'intégration éventuelles, l'unité en décibels ou en valeur efficace. Il va également s'agir de paramètres issus du contour temporel de l'énergie : coefficients de régression et erreur moyenne associée (R.Tato, 2002, A.Batliner 2000). Le shimmer est parfois inscrit dans les vecteurs acoustiques (N.Amir, 2001). Des mesures d'énergie à basse (< 250 Hz) et à haute fréquence (> 250 Hz) interviennent également (P-Y.Oudeyer 2002, V.Hozjan 2003, D.Ververidis 2004 et 2005, M.Shami 2007) ou seulement à basse fréquence (<500 Hz et <1000 Hz) complétées de valeurs RMS et des caractéristiques statistiques de la distribution des valeurs efficaces (T.Seppänen, 2003). Il peut enfin s'agir uniquement de valeurs RMS (J.Liscombe, 2003).

Descripteurs Temporels

Ils expriment les variations du débit phonatoire (K.Soltani, 2007) ou en sont indirectement rattachés par la mesure de diverses grandeurs comme la durée de phonèmes, de syllabes, de consonnes (V.Hozjan, 2003, J.Liscombe 2003) mais aussi la durée d'éléments sonores en unités de tranches de signal rapportée à la distance moyenne entre les vecteurs M.F.C.C des dites tranches (O-W.Kwon, 2003) ou encore les durées des parties audibles et inaudibles du signal qualifiées comme telles par rapport à un seuil d'énergie (S.Yacoub, 2003).

Sont repérés également le nombre et la durée des régions voisées et non voisées du signal (A.Batliner 2000, R.Tato 2002, H-Y.Cho 2004) complétés par la mesure des longueurs maximales des silences et de rapports entre la durée des silences et celle de la parole (T.Seppänen, 2003).

On peut noter enfin l'estimation des durées des « plateaux » du contour temporel de F_0 et de l'énergie (D.Ververidis, 2004 et 2005).

Descripteurs Spectraux

Il peut s'agir de la fréquence des premiers formants (associés à leurs statistiques respectives) mais aussi de l'énergie en deçà de 250 Hz (D.Ververidis, 2004), éventuellement complétés de la largeur de bande des formants et de critères relatifs à la distribution de l'énergie (R.Tato, 2002).

Cette distribution, lorsqu'elle est exprimée sous forme de calculs d'énergie par bandes critiques, fournit un vecteur acoustique de 12 valeurs qui permet d'aboutir à de meilleures performances qu'avec 12 coefficients cepstraux en échelle Mel (M.F.C.C) ou 16 coefficients cepstraux issus de la prédiction linéaire (L.P.C.C) : 77% en moyenne pour les différentes émotions contre 57% environ (T.L.Nwe, 2003).

La distribution spectrale de l'énergie est également présente pour la classification à partir d'un découpage fréquentiel en 18 bandes issu de transformations en ondelettes. Les auteurs proposent alors un vecteur acoustique composé d'une grandeur énergétique issue de ces bandes complété de mesures d'autocorrélation d'une tranche à l'autre (R.Sarikaya, 1998). Les performances seraient ainsi améliorées d'environ 7% par rapport au même système avec un vecteur M.F.C.C. Cette distribution est également prise en compte avec un banc de filtre

classique (125 Hz – 3.5 KHz) et avec un banc appliqué à l'enveloppe de Hilbert du signal pour prendre en compte les fréquences de modulation du signal. Sur un schéma représentant l'énergie obtenue pour les filtres de chaque type, on distingue des formes différentes suivant l'émotion simulée. Les vecteurs acoustiques qui en résultent conduisent à des taux de reconnaissance supérieurs à 90% (S.Wu, 2011).

La pente spectrale des voyelles apparaît également dans certains vecteurs acoustiques (J.Liscombe 2003). Elle est exprimée aussi sous la forme du rapport entre l'énergie dans la bande 2000 – 4000 Hz avec celle de l'énergie dans la bande 0 – 1000 Hz (H-Y Cho, 2004).

Autres Descripteurs

Comme il vient d'être mentionné, les vecteurs M.F.C.C apparaissent souvent parmi les descripteurs car ils sont à la base des systèmes de reconnaissance du locuteur. Mais ils sont le plus souvent complémentaires des descripteurs précédents. La reconnaissance peut également être fondée sur les seuls niveaux de sortie d'un banc de 13 filtres Mel (C.Busso, 2007) ou sur des coefficients cepstraux obtenus après un filtrage spécifique en sous-bandes (Y.Zhou, 2009). La moyenne et l'écart-type des coefficients M.F.C.C calculés à l'échelle du phonème (voyelle voisée ou non voisée) montrent de meilleurs taux de reconnaissance que ceux de vecteurs prosodiques calculés à l'échelle de la prononciation entière du mot. Les résultats sont meilleurs pour les consonnes que pour les voyelles pour une des deux bases de données vocales émotionnelles testées et très voisins pour la seconde (D.Bitouk, 2010).

Une approche originale consiste à extraire des sections de Poincaré du signal temporel des quatre premiers formants et de l'énergie, les caractéristiques statistiques suivantes : déterminant et trace d'une matrice de covariance des sections, coefficient de détermination et somme du carré des erreurs entre valeurs issues des plans des phases et celles issues d'une régression linéaire. En ajoutant ces grandeurs à celles de l'énergie, de la fréquence fondamentale et de durées des prononciations pour constituer un nouveau vecteur acoustique, le taux de reconnaissance a été accru d'environ 10% pour atteindre des scores moyens de l'ordre de 75% (K.Ślot, 2009).

Une autre approche originale consiste à décomposer les vecteurs acoustiques en une part propre à la fonction linguistique et en une part strictement responsable du caractère émotionnel de la parole. Il s'agit de séparer dans les contours temporels de F_0 la forme propre due à l'intonation et à l'accent en situation émotionnelle neutre du « bruit » provoqué par les effets de l'émotion sur ce contour. Une analyse discriminante procède à l'estimation du coefficient de proportionnalité entre le vecteur neutre et le vecteur émotionnel. La reconnaissance automatique est alors améliorée de plus d'une dizaine de pour cent pour atteindre des scores au delà de 90% (D-N.Jiang, 2004).

Enfin, en complément de descripteurs prosodiques, énergétiques et de durée, la prise en compte de la fonction de transfert du conduit vocal et plus précisément de la réponse en phase du filtre tout-pôle de l'analyse L.P.C a conduit à améliorer les scores de reconnaissance des émotions de plus de 10% (V.Sethu, 2007).

Il est difficile d'extraire une liste minimale de descripteurs commune aux études, aux bases de données et aux méthodes de classification.

Certains travaux opèrent avec un grand nombre de paramètres acoustiques qu'ils réduisent tout en maintenant les performances de reconnaissance : 276 descripteurs à 75 (B.Schuller, 2005), de 144 grandeurs à 26 (V.Hozjan, 2003) et de 87 à 10 (D.Ververidis, 2004) en notant pour ce dernier que les paramètres les plus pertinents sont la valeur maximale

de l'énergie, l'énergie en deçà de 250Hz normalisée par la longueur de la prononciation, et la valeur moyenne des pentes de croissance du contour de la fréquence fondamentale (D.Ververidis, 2005).

Pour d'autres auteurs les caractères acoustiques les plus pertinents sont différents : ce sont les vecteurs M.F.C.C avec des centroïdes (B.Schuller, 2005) et des caractéristiques statistiques d'autres descripteurs (R.Fernandez, 2005). Une étude relève que ce sont les données statistiques relatives à l'intensité qui sont les plus importantes suivies de celles de la fréquence fondamentale et que la prise en compte de l'énergie en deçà de 250 Hz améliore les scores de la reconnaissance pour plusieurs classificateurs (P-Y.Oudeyer, 2002).

Un travail de recherche conserve un très grand nombre de paramètres acoustiques (318) pour la reconnaissance (A.Shaukat, 2008). Le vecteur peut aussi être composé d'emblée d'une quinzaine de descripteurs (K.Soltami, 2007), de plus de cinquante (O.W.KWon, 2003) ou de trente deux décrits comme les marqueurs les plus robustes de l'émotion (S.McGilloway, 2000). Cette dernière étude montre que pour les huit paramètres acoustiques les plus pertinents, au moins un d'entre eux permet de distinguer les différentes émotions simulées lorsqu'elles sont considérées par paires.

Il est également observé une augmentation du taux de reconnaissance avec la diminution du nombre de paramètres acoustiques (de 16 à 8) (K.Soltani, 2007). Au contraire, le passage de un à dix descripteurs accroît nettement la performance (D.Ververidis, 2004 et 2005). Sur la base de 85 descripteurs statistiques (de la prosodie, des formants, de durées, du jitter, du shimmer) il a été démontré qu'au-delà de trente caractéristiques les performances de la reconnaissance ne sont pas améliorées et atteignent des scores voisins de 80% (H.Hu, 2007).

Une étude montre que quel que soit le nombre maximal de descripteurs considéré (de 36 à 166 de tout type) les scores présentent un maximum pour un nombre de paramètres différent du total (56 pour un total de 166 par exemple). En outre, plus le nombre total de grandeurs est important, meilleure est la reconnaissance pour le système utilisé par les auteurs (A.Hassan, 2009). Des algorithmes spécifiques sont développés pour procéder au choix du sous-ensemble de descripteurs qui parmi l'ensemble de départ fournira les meilleurs résultats (A.Álvarez 2007, A.Hassan 2009, D.Hübner 2010).

On note enfin, qu'en combinant les scores d'un procédé SVM de reconnaissance où les descripteurs sont essentiellement prosodiques avec ceux d'un procédé GMM basé sur des vecteurs M.F.C.C, le taux de reconnaissance des émotions est accru (H.Hu, 2007). Le même type de combinaison entre classificateurs pour des descripteurs prosodiques et spectraux améliore la reconnaissance. La fusion des performances des deux co-classificateurs opère sur des paramètres sélectionnés comme les plus pertinents après application d'un critère fondé sur les distances inter et intra – classes. Ce sont les paramètres statistiques de l'énergie, de la fréquence fondamentale, de sa pente pour la dernière voyelle prononcée et de la durée des voyelles (I.Luengo, 2009).

D'autres méthodes combinatoires entre systèmes de classification existent (I.Lopez-Moreno, 2009, L.Xu 2009). Elles contribuent toutes à améliorer les performances même si les types de descripteurs ne sont pas toujours les mêmes. La confrontation de plusieurs types de classificateurs a également été mise en place sur des corpus identiques pour en comparer les performances (F.Dellaert 1996, N.Amir 2001, P-Y.Oudeyer 2002).

On peut également noter la réalisation de bonnes performances (taux supérieurs à 65%) avec comme seul descripteur l'évolution temporelle de F_0 (H.Sato, 2001).

Enfin, selon le type d'émotion on peut distinguer des ensembles de descripteurs ou des paramètres uniques pour lesquels sont obtenues les meilleures performances. Par exemple, un score de 81,25% avec la seule valeur maximale de F_0 pour reconnaitre la tristesse et un score de 80,11% avec la pente spectrale d'une voyelle pour l'ennui (J.Liscombe, 2003). La colère et la tristesse apparaissent comme plus faciles à détecter que l'ennui et le bonheur (H-Y.Cho, 2004).

Toutes ces observations illustrent la difficulté de synthèse bibliographique.

Néanmoins, quel que soit le type de système et de langue, il apparaît que le squelette minimal du vecteur acoustique nécessaire à une reconnaissance de bonne qualité des émotions simulées doit comprendre au moins quatre caractéristiques respectivement liées à la fréquence fondamentale, à l'énergie, à la durée et à un contour temporel toutes accompagnées de quelques valeurs statistiques associées et éventuellement complété de quelques descripteurs spectraux. Ces derniers accroissent la performance par rapport aux seuls paramètres prosodiques (D.Bitouk, 2009).

La plupart des études renvoient au même schéma : un ou plusieurs classificateurs, une base de données acoustique d'émotions simulées à la fois pour l'entrainement et le test des systèmes, une liste des descripteurs utilisés et les performances obtenues. Ces dernières varient suivant la base de données, l'émotion considérée et les méthodes de reconnaissance employées. Les taux de reconnaissance sont supérieurs à 60% et peuvent être bien plus élevés pour certaines expérimentations, pouvant dépasser 80% pour certaines émotions simulées comme la colère, l'état de neutralité (K.Soltani, 2007), 90% pour la joie (H.Sato, 2001), 80% pour la tristesse (D.Ververidis, 2004). La « même » joie peut être reconnue à environ 40% pour d'autres travaux (V.Hozjan, 2003). D'une étude à l'autre les scores diffèrent pour le même type d'émotion.

Le niveau d'activation de l'émotion (« arousal ») a été lui aussi soumis à la reconnaissance automatique (élevé, neutre et faible) avec des vecteurs combinant prosodie, énergie, durée et spectre pour aboutir à un score moyen de 82% (59% indépendamment du locuteur) (R.Tato, 2002).

Il est observé des performances similaires et élevées pour des vecteurs et des méthodes de classification sensiblement différents. Il est également observé des scores parfois très différents d'une émotion à l'autre pour différentes études. Selon les algorithmes d'apprentissage et de classification, les taux moyens de reconnaissance peuvent varier de plus de 10% pour les mêmes bases de données. Mais leur performance individuelle ne conduit pas à les situer au même rang selon la base de données (M.Shami, 2007).

Il semble en revanche que les calculs des descripteurs sur la base de segments de parole ou sur des prononciations plus longues aboutissent à des performances voisines (M.Shami, 2007) sauf dans le cas de systèmes de classification HMM où le type de phonème pris en compte modifie sensiblement les scores (voyelle, nasale, fricative, tous phonèmes).

Des auteurs concluent que les voyelles sont les plus importants indicateurs de l'émotion en termes de reconnaissance automatique (C.M.Lee, 2004). L'extraction de caractéristiques acoustiques au niveau du phonème parait être très pertinente en vue de la reconnaissance émotionnelle. Les performances sont meilleures pour des descripteurs spectraux mesurés au niveau du phonème plutôt qu'au niveau de la prononciation complète (D.Bitouk, 2009 et 2010). Ce type d'analyse a été réalisé conjointement sur des voyelles et

des consonnes et montre de meilleurs résultats que celle de l'approche traditionnelle réalisée sur des segments voisés – non voisés (F.Ringeval, 2008). De plus, il apparaît que la reconnaissance de différentes émotions est mieux réalisée pour certains phonèmes que pour d'autres : [a] par exemple (V.Sethu, 2008).

Par ailleurs, il semble que les effets acoustiques et articulatoires de l'émotion soient plus significatifs en milieu de phrase que pour les mots prononcés à son début et à sa fin car les scores de la reconnaissance y sont sensiblement plus élevés (J.Kim, 2009).

La conclusion des résultats en matière de reconnaissance d'émotions simulées, et qui sera confirmée pour les autres types d'émotion présente plusieurs points :

- les recherches en la matière ont conduit à proposer une large gamme de descripteurs classiques et originaux nourrissant ainsi l'analyse acoustique ; mais relativement peu d'information existe sur les conditions de la mesure ;
- l'association de ces paramètres pour former des vecteurs servant à la classification des émotions masque leurs variations individuelles ou conjointes. Cette information disparait au profit d'une estimation globale multidimensionnelle ;
- le type de caractéristique le plus approprié n'a pas encore été trouvé (M.Kockmann, 2009) surtout pour des applications où l'émotion qui se manifeste ne couvre que quelques mots. Les bases de données ou les expérimentations spécifiques pour simuler l'émotion traitent nécessairement de grands corpus avec une approche globale ;
- le nombre optimal de descripteurs et leur type ne sont pas connus. Une redondance entre les multiples descripteurs existe mais elle n'est pas explicitée ;
- le « poids » de chaque paramètre acoustique, ou de groupes de descripteurs, dans la reconnaissance globale n'est pas quantifié ;
- les vecteurs utilisés et les classificateurs diffèrent, si bien que les comparaisons sont difficiles même si des tentatives pour améliorer la compatibilité des résultats sont mises en œuvre (B.Schuller, 2009) ;
- on ne dispose pas de règles entre les valeurs numériques des paramètres (arbres de décision, seuils) pour identifier le profil acoustique des émotions.

2.2.1.2. Émotions Induites par le Contexte Expérimental

Dans ce paragraphe vont être répertoriées des études où il ne s'agit plus de simuler des émotions, mais d'expériences dont le protocole va provoquer des réactions émotionnelles chez le locuteur. L'intérêt principal des ces situations réside dans le fait que les échanges verbaux ont un caractère plus naturel puisqu'il n'est pas demandé de jouer une émotion. En contrepartie, l'expression émotionnelle vocale est sans doute plus subtile et probablement sujette à de plus grandes variabilités.

Il peut s'agir :

- de provoquer le sourire des locuteurs (C.Émond 2006, A.Drahota 2008) ;
- de demander aux sujets de parler d'expériences personnelles évoquant différentes émotions (joie, peur, tristesse …) (B.Zei 1998, N.Amir 1998 et 2000, D.Erickson 2004, D.Rochman 2008) ;

- de susciter une prise de parole en public (A.M.Goberman, 2011) ;
- d'utiliser un protocole où les locuteurs communiquent avec un ordinateur dont on programme des dysfonctionnements et on manipule ses performances afin de modifier l'attitude du sujet (A.Batliner 2000 et 2003, V.Aubergé 2004, S.Rossato 2004) ;

Une autre application consiste à enregistrer les conversations d'enfants avec un robot (« Aibo » de Sony) qu'ils contrôlent vocalement mais qui va parfois leur désobéir pour provoquer des réactions émotionnelles (B.Schuller 2007, D.Seppi 2008, C-C.Lee 2009, E.Bozkurt 2009, M.Lugger 2009, T.Polzehl 2009, P.Dumouchel 2009, T.Vogt 2009, R.Barra-Chicote 2009). Cette dernière situation a constitué la banque de données fournie aux chercheurs pour tester les systèmes de reconnaissance automatique lors du congrès Interspeech 2009 pour le « Emotion Recognition Challenge ».

Dans des contextes similaires, un tuteur adulte joue le même rôle que l'ordinateur (T.Zhang 2004, D.Litman 2006). Il existe aussi une étude où les sujets perçoivent des gains selon l'exactitude de leurs réponses dans le cadre d'un scénario contrôlé pour induire des réponses verbales au contenu émotionnel (J.Kim 2005). Un protocole semblable a été mis en place avec des questions de Physique à des étudiants (H.Ai, 2006). Le même type de script existe aussi pour simuler par téléphone des interactions avec une plateforme vocale de réservation ou de service aux clients (R.Kumar 2006, T.Arimoto 2007). Un jeu en ligne (Y.Arimoto 2008) ou des jeux vidéo (T.Johnstone 2005, K.P.Truong 2009) ont aussi servi de support expérimental pour provoquer de l'émotion.

Ces travaux sont plus récents que ceux sur la simulation d'émotion à l'origine des recherches en la matière. Dès lors, les études traitent surtout de reconnaissance automatique des émotions provoquées et peu sont exclusivement consacrées à l'analyse acoustique. Néanmoins, un volet analyse figure toujours avec la présentation des descripteurs acoustiques.

Pour l'essentiel, ces paramètres sont les mêmes que ceux qui ont été choisis pour la reconnaissance automatique des émotions simulées. On note deux études où plus de 4000 paramètres acoustiques ont été mis en œuvre (B.Schuller 2007, D.Seppi 2008). Des caractéristiques acoustiques originales ont été introduites.

Il s'agit de gradients spectraux qui mesurent la pente spectrale entre l'amplitude de la fréquence fondamentale et celle de chacun de ses vingt premiers harmoniques (M.Lugger, 2009). C'est aussi le cas d'un histogramme en demi-tons de l'évolution du contour temporel de la fréquence fondamentale normalisée à l'octave et à partir duquel sont calculées des fonctions d'autocorrélation. Leur combinaison linéaire pondérée permet de définir un paramètre de dissonance (M.Lugger, 2009).

Enfin, des mesures physiologiques et acoustiques ont conduit à élaborer des ensembles de descripteurs propres à chaque type et un ensemble commun pour indiquer que leur complémentarité accroit sensiblement les performances de la reconnaissance. Les seuls éléments physiologiques (électromyogramme, électrocardiogramme, conductivité de la peau, modifications de la respiration) conduisent à des scores similaires à ceux des seules mesures acoustiques (J.Kim, 2005). Ce dernier point contribue à témoigner de la pertinence des mesures acoustiques non invasives pour la détection des perturbations émotionnelles.

En ce qui concerne la méthode pour discriminer les émotions, des indices émotionnels sont construits sur la base de calcul de distances normalisées entre vecteurs acoustiques essentiellement prosodiques, énergétiques et temporels (A.Amir, 1998 et 2000). À noter que

les locuteurs qui sont retenus possèdent des mesures physiologiques (électromyographie, rythme cardiaque et réponse galvanique de la peau) dont les variations restent comprises dans un intervalle de confiance fixé au préalable par les expérimentateurs.

La classification automatique peut être précédée d'une « labellisation », par exemple par des auditeurs spécialisés en linguistique, pour former des ensembles de prononciations identifiées comme proche d'une manifestation émotionnelle donnée (B.Schuller 2007, D.Seppi 2008) ou, par le locuteur lui-même (V.Aubergé 2004, K.P.Truong 2009). Celui-ci n'ayant pas à simuler une émotion particulière, il convient donc de pré - établir le classement sans quoi les performances sur la base brute des données risquent d'être faibles (B.Schuller 2007, D.Seppi 2008).

Une autre approche consiste à effectuer des étapes successives de classements binaires depuis le caractère émotionnel ou non jusqu'aux catégories d'émotions (C-C.Lee, 2009). Un système de reconnaissance automatique de la parole peut être utilisé pour détecter des éléments lexicaux spécifiques associés à l'expression émotionnelle en complément des descripteurs classiques utilisés par un système de reconnaissance de l'émotion (T.Zhang, 2004).

Le classement effectué n'est donc pas toujours celui d'émotions bien déterminées comme pour la simulation par des acteurs. Il est dépendant du protocole expérimental mis en place pour provoquer les réactions émotionnelles.

Dans le cas de dialogues spontanés avec un tuteur, les catégories émotionnelles sont la confusion, la clarté, l'incertitude (R.Kumar, 2006) ; pour un scénario avec des enfants on va distinguer la perplexité, la confiance et l'hésitation (T.Zhang, 2004) ; pour une étude où les locuteurs doivent s'exprimer en se remémorant des situations personnelles les catégories peuvent être celles des émotions primaires (N.Amir, 2000) ; enfin dans le cas d'un jeu vidéo, on trouve des classes d'émotion comme l'émerveillement, le malin plaisir ou l'excitation (K.P.Truong, 2009).

En termes de résultats, la performance chute lorsqu'il s'agit d'une émotion provoquée par un scénario par rapport au même type d'émotion jouée par un acteur (A.Batliner, 2000). Les scores varient également suivant la « labelisation » effectuée (D.Seppi, 2008). D'un point de vue général, les performances obtenues sont moins bonnes que lorsque les systèmes opèrent sur des corpus d'émotions jouées par les locuteurs et ce, avec les mêmes stratégies de classification. Les scores moyens ne dépassent pas une cinquantaine de pour cent alors qu'ils peuvent atteindre 80% et plus pour les émotions jouées.

Les analyses acoustiques sur les corpus de parole émotionnelle non simulée sont peu nombreuses. On relève toutefois une étude des contours temporels de la fréquence fondamentale mettant en évidence des différences de forme et de dynamique par rapport à ceux issus d'une émotion jouée (V.Aubergé, 2004) et des travaux ciblant les liens de la prosodie avec le contexte sémantique du discours (R.Kehrein, 2002).

L'émotion spontanée se traduit également par des différences en termes d'articulation (mouvement et position de la langue) par rapport à celle qui est simulée (D.Erickson, 2004).

La parole souriante induite par un protocole expérimental montre que l'étendue fréquentielle de sa fréquence fondamentale est plus petite que pour la parole normale (C.Émond, 2006). L'auto-évaluation et l'estimation par des auditeurs de l'anxiété provoquée par la prise de parole en public sont liées au pourcentage de pauses dans le discours et à l'écart-type de la fréquence fondamentale moyenne (A.M.Goberman, 2011).

La durée des prononciations et la pente spectrale apparaissent comme plus appropriées à caractériser la colère que les paramètres prosodiques. Le résultat est obtenu en comparant des régressions linéaires entre les descripteurs et des notes d'appréciation subjective par des auditeurs (Y.Arimoto, 2007). Le même auteur effectue une analyse de variance qui lui permet de conclure aux émotions qui sont le mieux discriminées par les descripteurs classiques qu'il utilise dans le contexte expérimental d'un jeu en ligne (Y.Arimoto, 2008).

Lors d'analyses conjointes de la variabilité de la fréquence cardiaque et de paramètres acoustiques (F_0, niveau sonore, débit phonatoire) de locuteurs devant raconter des situations émotionnelles réellement vécues, il a été montré que ce paramètre physiologique était en lien avec la capacité individuelle à différencier vocalement les états affectifs (B.Zei, 1998).

Enfin, une expérience suscitant de l'émotion grâce à l'utilisation d'un jeu sur ordinateur a montré que le niveau sonore moyen tout comme la fréquence fondamentale et la durée des prononciations pouvaient varier significativement (T.Johnstone, 2005).

L'analyse et la reconnaissance des émotions suscitées par le contexte expérimental s'avère plus difficile que dans le cas des émotions simulées par des acteurs. La nature de l'épisode émotionnel est moins clairement identifiable tout comme sa durée et ses instants de commencement et de fin. Ceci est particulièrement vérifié lorsqu'il s'agit de procéder à une pré-classification auditive des prononciations pour la reconnaissance. Les performances des systèmes sont moins élevées.

En matière d'analyse, les descripteurs sont identiques à ceux examinés pour la simulation d'émotion, confirmant leur sensibilité commune et révélant l'absence d'indicateur ou de domaine de variation propres aux deux types de conditions expérimentales.

2.2.1.3. Perception des Émotions

Un auditeur peut percevoir l'état émotionnel d'un locuteur à partir de son expression vocale (R.Banse, 1996).

La plupart des travaux de reconnaissance des émotions comprennent un volet perceptif. La liste bibliographique reprend donc celle des publications sur la reconnaissance mentionnées dans les paragraphes précédents.

Il s'agit, notamment pour les émotions simulées, de s'assurer de la pertinence du corpus de parole soumis à reconnaissance automatique. Pour cela, les auteurs se livrent à des tests où des auditeurs ont pour tâche de reconnaître et de classer les émotions. Il existe une station automatisée adaptée aux tests de perception et d'évaluation auditive (A.Ghio, 2003) et un instrument où l'auditeur pointe à l'intérieur d'un cercle sa perception de l'état émotionnel selon la dimension de valence et d'activation (R.Cowie, 2000).

Toutes les études mentionnent des scores élevés. La reconnaissance auditive des émotions offre toujours de meilleurs résultats que les systèmes informatiques (A.Shaukat, 2008). Elle est également appliquée pour l'analyse du signal où elle permet de valider des adjectifs descripteurs de l'émotion jouée, aboutissant à la conclusion que les paramètres prosodiques sont fortement corrélés avec ces attributs perceptifs (T.Ehrette, 2002). Elle peut être aussi mise en œuvre pour valider le contenu expressif de contenus publicitaires pour être utilisés en synthèse sonore (I.Iriondo, 2007).

Le même type de vérification ou de validation a été effectué pour les bases de données vocales d'émotions simulées qui servent à analyser les signaux et à tester les systèmes de reconnaissance automatique (R.S.Bolia, 2003 pour la base SUSAS). Si la reconnaissance est

effective dans les cas d'émotions simulées, on note que pour un corpus de conversation naturelle de la vie de tous les jours, les labels émotionnels classiques (joie, surprise, doute, dégout …) qui sont proposés aux auditeurs ne suffisent pas à différencier les émotions (N.Campbell, 2004).

La synthèse qui suit présente des travaux spécifiques à la perception des émotions. Les résultats présentent un intérêt pour l'analyse car ils témoignent de l'existence des modifications acoustiques du signal. On se place ici hors du contexte sémantique, étant entendu que le sens des mots prononcés peut fortement participer à la reconnaissance auditive d'une émotion.

Des relations significatives entre les aspects sémantiques et les caractéristiques acoustiques ont été mises en évidence (C-F.Huang, 2008). La syntaxe peut elle aussi jouer un rôle car le souvenir d'évènements est raconté différemment selon l'émotion qui lui est associée s'agissant par exemple du nombre de propositions ou du nombre moyen de mots par proposition (T.Habermas, 2009). À la seule lecture de la transcription de dialogues de films, la reconnaissance de la tristesse et de la colère s'est avérée 15% moins précise qu'à l'écoute de la bande son correspondante (T.S.Polzin, 2000).

Comme pour l'analyse acoustique dont les résultats s'inscrivent dans un champ inter – culturel, des auditeurs peuvent détecter les émotions exprimées dans une autre langue que la leur (A.Abelin 2000, K.R.Scherer 2000). Il est également observé la faculté d'auditeurs à discriminer des émotions jouées d'émotions spontanées (N.Audibert, 2008).

La perception de trois dimensions émotionnelles que sont le degré d'activation (« arousal »), le plaisir (lié à la « valence ») et l'intensité (« power ») est également possible pour plusieurs émotions primaires (C.Pereira, 2000) ou dans le cadre de conversations (J.C.Acosta, 2009). En demandant à des acteurs de simuler des émotions avec différents types d'intonation, la catégorisation des auditeurs a conduit à établir que le contour temporel est plus décisif pour reconnaitre l'émotion jouée que la fréquence fondamentale moyenne (E.Rodero, 2011). La reconnaissance auditive du caractère positif ou négatif (valence) d'émotions a également montré que l'appréciation des auditrices diffère de celle des auditeurs en termes de dispersion des scores (plus élevé pour ces dernières) (M.Swerts, 2008). La discrimination positive – négative est en outre plus marquée pour l'écoute d'acteurs professionnels que d'acteurs inexpérimentés (E.Krahmer, 2008). Le couplage de tests perceptifs et de mesures acoustiques suggère qu'il est possible d'identifier la valence à partir de la durée des voyelles (plus longue pour les émotions positives) (T.Waaramaa, 2010).

Comme précédemment, les écoutes comparatives de plusieurs enregistrements d'acteurs simulant des émotions a permis de montrer que le contour temporel de la fréquence fondamentale F_0 contribuait plus que la fréquence elle-même à la reconnaissance des émotions (C.Pereira, 2010). Le même type d'investigation accompagné de mesures acoustiques du contour de F_0, indique que l'émotion peut être reconnue par des auditeurs même si les acteurs ne prononcent qu'un seul mot (I.Linnankoski, 1997 et 2005) voire monosyllabique (Y.Nagasaki, 2004). Ces deux derniers auteurs établissent des propriétés acoustiques des émotions qu'ils ont étudiées à partir de mesures de F_0, de durées et de forme spectrale.

Les tests d'écoute ont également permis de conclure que des onomatopées et des interjections verbales ou non verbales possèdent un contenu émotionnel reconnu et classé par les auditeurs. Ces éléments participent donc à la communication des émotions (M.Schröder 2000 et 2003, C.T.Ishi 2008, E.R.Simon-Thomas 2009).

De même, la contribution sonore à la perception audiovisuelle du sourire est mise en évidence (V.Aubergé 2000 et 2003, A.Drahota 2008). Les auteurs suggèrent que la pente de décroissance de F_0 est le principal paramètre modifié par le locuteur pour exprimer cette émotion (V.Aubergé 2000 et 2003). D'autres caractéristiques semblent elles aussi jouer ce rôle comme la fréquence des deux premiers formants (A.Drahota, 2008).

La perception auditive de l'émotion existe, comme en témoigne cette revue de la littérature, mais elle ne peut constituer une preuve irréfutable de son existence ou de son absence. Des jurys d'auditeurs peuvent être amenés à se prononcer non seulement sur sa présence mais aussi sur sa nature et son degré suivant les protocoles mis en place. Pour autant, l'absence de résultats significatifs ne conduit pas à rejeter l'hypothèse de son existence. L'existence de micro-tremblements des cordes vocales, non audibles, en est une preuve (cf § 2.4.1 page 43).

Un contrôle conscient de l'état émotionnel peut éliminer les traces audibles de l'émotion intérieure. L'analyse directe du signal, ou indirecte pour la reconnaissance automatique, n'est donc pas dépendante de la perception auditive. Des résultats positifs peuvent par contre être un précieux indicateur pour l'expérimentateur analysant le signal, ne serait-ce que pour en favoriser la localisation dans un discours naturel et continu.

2.2.1.4. Émotions Simulées par Synthèse Vocale

L'examen de la littérature qui traite de la synthèse vocale pour simuler une parole émotionnelle présente l'intérêt de fournir les paramètres acoustiques sur lesquels il faut agir, de quelle sorte et avec quel degré de traitement.

Dès les années soixante des manipulations du signal synthétique (fréquence fondamentale, enveloppe) sont effectuées pour en estimer les conséquences sur la reconnaissance auditive des émotions (P.Lieberman, 1962). L'intelligibilité de la parole synthétique étant maitrisée, l'application principale des recherches est d'éliminer son caractère mécanique et de la rendre plus naturelle (I.R.Murray, 1996). Des logiciels de traitement du signal destinés à des applications audiovisuelles proposent déjà des fonctions pré-réglées pour transformer par exemple une voix en voix stressée, mais les résultats ne sont pas encore probants.

En regard des manipulations du signal, figurent les tests perceptifs qui permettent de valider la qualité de la synthèse. La question qui se pose ici est d'identifier les paramètres à modifier pour qu'une parole synthétique possède un caractère émotionnel. Ils ont été classés en trois types (et décrits pour les émotions primaires et secondaires) : la qualité de la voix, le rythme des prononciations et le contour temporel de la fréquence fondamentale (I.R.Murray, 1993).

Les expérimentateurs opèrent des modifications sur ces paramètres, et d'autres, pour soumettre le signal modifié à des auditeurs et conclure sur leur influence dans la catégorisation. La base est soit l'enregistrement d'un discours neutre soit celui que génère un synthétiseur à partir d'un texte écrit.

À la lueur des analyses acoustiques et des performances des descripteurs pour la reconnaissance automatique de l'émotion, c'est sans surprise que les paramètres prosodiques que sont la fréquence fondamentale, son contour temporel, le niveau sonore et les durées des structures du langage apparaissent comme les principales grandeurs utilisées.

Les manipulations de ces grandeurs se cumulent à leurs variations naturelles dues à l'intonation ainsi qu'aux traits distinctifs qui peuvent permettre d'identifier le sexe du

locuteur ou d'estimer son âge. Concernant l'intonation et l'émotion, la première participe à l'expression sonore de la seconde (T.Bänziger 2005, I.Grichkovtsova 2007 et 2008).

Les études qui attestent, par des tests d'écoute, de la capacité des caractéristiques prosodiques à habiller la parole synthétique d'un aspect émotionnel sont nombreuses. Elles procèdent à diverses modifications du signal issu de prononciations d'au moins plusieurs mots.

En appliquant les valeurs de plusieurs paramètres des contours temporels de F_0 (attaque, décroissance, valeur moyenne, gamme de variation) et de l'intensité, mesurés pour diverses émotions simulées, à un discours neutre monosyllabique, il est montré que des émotions peuvent être reconnues, mais inégalement, avec des confusions entre les émotions (N.Audibert, 2005). Dans une étude plus récente, les auteurs ont montré qu'une identification précoce des émotions est possible dès le début de la structure monosyllabique (N.Audibert, 2007).

Sur des prononciations plus longues, les modifications du contour temporel de la fréquence fondamentale jouent aussi un rôle pour allouer un caractère émotionnel au discours (D.R.Ladd 1985, S.J.L.Mozziconacci 2000, K.Hirose 2004).

La méthode la plus courante pour transformer les contours consiste à les modéliser par des segments successifs de droite auxquels sont affectés des paramètres (pente par exemple) qu'il est ensuite possible de modifier individuellement (S.J.L.Mozziconacci 2000, V.Sethu 2009). D'autres types de transformation ont été envisagés comme par exemple de relever ou d'abaisser le contour uniquement en fin de phrase ce qui en a montré le lien avec la perception de la valence des émotions : tendance croissante pour les émotions positives et décroissante pour les négatives. L'aspect sémantique joue un rôle important dans la perception puisque les prononciations filtrées (bande passante 500 – 1000 Hz) rendues ainsi non compréhensibles sont moins bien discriminées selon leur valence (H.Yamasaki, 2004).

La génération de contours temporels émotionnels à partir de texte s'accompagne également de règles pour inclure l'accent et prédire la durée des segments de parole (K.Hirose 2004).

Un modèle à deux masses des cordes vocales prend en compte les modifications de facteurs de perturbations de la fréquence fondamentale et de l'amplitude (« Pitch and Amplitude Perturbation Quotient » PPQ et APQ) mesurés pour différentes émotions afin de synthétiser l'expression émotionnelle de bébés (C.Araki, 2008).

En minimisant les différences entre les représentations temps-fréquence d'une parole neutre avec celles de mesures effectuées sur des signaux de parole émotionnelle, il a été possible de reproduire l'émotion (M.Kobayashi, 2005).

Par une méthode de transplantation (« paradigme de transplantation ») et d'échange du contenu prosodique entre deux prononciations où le contour prosodique d'une émotion jouée est reporté sur le même discours synthétique neutre à partir du texte et inversement, il est possible d'estimer les effets respectifs de la prosodie et de la qualité de la voix sur la perception des émotions. Les résultats montrent que leur influence respective peut être différente suivant l'émotion considérée (I.Grichkovtsova, 2007 et 2008).

L'écoute de diverses transplantations indique que le contour temporel de F_0 participe activement à la reconnaissance des émotions positives alors que la durée et la qualité de la voix interviennent plus sur les émotions négatives (N.Audibert, 2006).

La méthode appliquée à l'identification de l'émotion sur des segments successifs de phrases (points d'ancrage) confirme aussi le fait que la détection auditive peut avoir lieu dès le début des structures phonétiques (I.Grichkovtsova, 2008).

Le même type de copie du contour de F_0 et des durées des structures vocaliques sur un signal neutre n'a que peu altéré les scores de l'identification auditive des émotions jouées (J.Vroomen, 1993). Ce transport s'est en revanche révélé démonstratif en utilisant les rapports de F_0, d'énergie et de durées entre les prononciations de plusieurs phrases neutre et celles prononcées en simulant des émotions (S-J.Kim, 2004).

Une méthode d'évaluation interactive a été développée pour que l'auditeur soit en situation d'entendre les conséquences de la modification de certains paramètres (intensité, F_0, durée) pour parvenir à la meilleure synthèse du type d'émotion proposé (C.Ovesdotter, 2006).

D'autres méthodes de conversion des contours temporels existent (J.Tao, 2006), la liste précédente n'est pas exhaustive. Mais quelles qu'elles soient, il apparait que les paramètres prosodiques influent fortement sur le caractère émotionnel de la parole synthétique.

La durée et l'amplitude des structures sonores de la parole y participent également. Le contrôle de ces caractéristiques (amplitude et durée des syllabes) a pu en montrer l'intérêt pour la qualité de la synthèse (K.Koike, 1998) en complément de modifications de F_0 (P-Y.Oudeyer, 2002).

La modification de la durée de prononciation de signaux de parole émotionnelle simulée semble permettre une meilleure catégorisation par l'auditeur que les manipulations des variations de F_0 (C.Breitenstein, 2001).

Avec la transposition sur un discours neutre du contour de F_0 pour différentes émotions, et en procédant à des séries de compressions-extensions du signal, des tests perceptifs montrent que les variations du débit phonatoire global sur la prononciation entière et sur des éléments constitutifs participent aussi au transport de l'émotion (S.J.L.Mozziconacci, 2000).

Tous ces paramètres (fréquence fondamentale, débit, niveau sonore) peuvent être contrôlés, par exemple par l'attribution de plusieurs niveaux pré-établis (J.Trouvain, 2006) ou appliqués à partir d'une base de données de parole émotionnelle (caractéristiques prosodiques) (P.Boula de Mareüil, 2002).

Outre les paramètres de la prosodie, le spectre joue aussi un rôle dans l'identification de la parole émotionnelle synthétique. Des études se sont attachées à le prendre en considération pour la synthèse des émotions, le plus souvent en complément des descripteurs prosodiques.

Les fréquences de raies spectrales issues d'une analyse L.P.C du signal de voyelle en situation émotionnelle sont appliquées au signal neutre pour réaliser la conversion vers une parole émotionnelle. Tous les autres paramètres étant par ailleurs identiques, les auditeurs préfèrent les versions sonores qui possèdent un spectre modifié pour la colère et la tristesse mais pas pour l'expression de la surprise (Z.Inanoglu, 2009).

À partir de prononciations modifiées par une technique de « morphing » où le spectre est manipulé par l'expérimentateur, l'analyse acoustique de celles qui ont fourni la meilleure reconnaissance auditive de la tristesse suggère que les formants sont en relation avec la perception de la tristesse (D.Erickson 2008, H.Kawahara 2003).

Un travail sur l'enveloppe spectrale, transposée en utilisant les coefficients M.F.C.C pour être synthétisée, montre qu'elle contribue peu à l'identification auditive de l'émotion

(R.Barra, 2007). L'application de modifications prosodiques et de l'enveloppe spectrale révèle que lorsqu'elles sont individuellement effectuées au niveau du phonème, les secondes sont plus efficaces pour la reconnaissance des émotions étudiées (M.Bulut, 2005).

Enfin, la « cartographie » de propriétés spectrales de diverses émotions (pentes spectrales, rapports d'énergie) permet d'introduire les modifications à apporter à un spectre neutre pour conférer à la synthèse un caractère émotionnel (C.Drioli, 2003).

Comme il vient d'être vu, l'essentiel des travaux de synthèse s'attache à produire une parole dont le contenu émotionnel est proche de celui d'émotions bien caractérisées (joie, peur …). Mais il ne semble encore pas possible d'agir pour créer des contenus plus subtils comme on les rencontre dans la vie de tous les jours. D'ailleurs, la littérature fournit peu de règles ou de méthodes précises à appliquer sur les paramètres du signal pour colorer émotionnellement la parole synthétique. La cause est certainement liée à l'analyse acoustique de la parole qui ne fournit pas précisément les données nécessaires à transposer dans les systèmes de synthèse pour créer les traits fins de l'émotion. Néanmoins, cet examen de la littérature corrobore l'identification et la sensibilité des paramètres acoustiques que l'analyse a mis en évidence.

2.2.2. Émotions Réelles

Il ne s'agit plus de simuler ou de provoquer des émotions pour en effectuer une analyse acoustique, une reconnaissance automatique ou à les créer dans la parole synthétique. Le propos est d'analyser le signal ou de procéder à une reconnaissance lors de situations réelles. L'expérimentateur n'agit donc pas sur la situation qui provoque l'émotion et ne la crée pas artificiellement. Au plus, il se contente de mettre en œuvre les moyens nécessaires à l'enregistrement des voix et à choisir les éléments de vocabulaire à traiter.

Les termes de stress ou de stress émotionnel sont également employés en lieu et place de l'émotion. Les situations réelles se prêtent difficilement à une typologie précise des perturbations provoquées. Sauf cas particulier, on identifiera ces dernières par la description des conditions qui ont conduit à les induire.

Les études qui s'inscrivent dans ce paragraphe traitent de situations, hors du contexte des transports, où les enregistrements ne sont donc pas initialement destinés à l'étude des émotions.

Deux familles de travaux peuvent être distinguées : celle où les émotions réelles sont capturées dans des programmes audiovisuels existants et celle où elles sont saisies parmi des échanges téléphoniques, par exemple vers des centres d'appels.

Hors de ces situations, on peut citer la capture et l'analyse de voix de bébés que le contexte a pu qualifier d'heureuses, d'impatientes, en colère et avec de la peur. Une classification et une reconnaissance fondées sur des descripteurs issus de l'opérateur de Teager et appliquée, non plus dans le domaine temporel mais dans le domaine fréquentiel, ont abouti à des scores particulièrement élevés (au-delà de 90%) (G.Hui, 2008).

Sources audiovisuelles d'émotions réelles **:** Il s'agit par exemple de locutrices fondant en larmes en racontant leurs expériences douloureuses lors d'une émission télévisée.

Deux catégories de stimuli (neutre et émotionnel) ont été proposés à des auditeurs en charge d'identifier ou pas la nature émotionnelle des extraites de parole.

Les résultats des tests auditifs montrent que l'émotion est mieux perçue dans la partie finale de la phrase que dans les autres parties sélectionnées. Les émotions sont traduites par une augmentation de la fréquence fondamentale moyenne, un ralentissement du débit et une augmentation du jitter. La durée moyenne de 100 ms (voyelles de l'expérience) semble être la durée minimale pour une perception de l'émotion (S-J.Chung, 1998).

Une autre étude sur des extraits de programmes radiophoniques et télévisuels a servi de base à une méthode d'annotation des données audio pour une extraction automatique des caractéristiques acoustiques (R.Stibbard, 2000).

La base de données « Belfast Naturalistic Emotion Database » possède également des données audiovisuelles de ce type. Les analyses effectuées ont montré que l'expression de la dimension activation de l'émotion s'accompagne d'un accroissement de la fréquence fondamentale et de sa gamme de variation, de phrases plus longues, de pauses plus courtes, de plus importantes et rapides variations de F_0, d'une augmentation de l'intensité et d'une pente spectrale moins importante (M.Schröder, 2001).

Une autre base de données « EmoTV » avec un corpus en français existe également (E.Douglas-Cowie, 2005). On peut citer aussi la base de données S.A.F.E (« Situation Analysis in a Fictional and Emotional database ») de séquences audiovisuelles extraites de films (C.Clavel, 2006 et 2008).

Des émissions télévisuelles politiques ont également servi de base à l'identification des caractéristiques acoustiques de la colère : augmentation de la fréquence fondamentale moyenne, diminution de la gamme de variation de F_0, diminution de la pente spectrale (N.Amir, 2003). La reconnaissance automatique d'émotions, à la fois sur un corpus joué et sur un corpus d'extraits télévisuels, avec des descripteurs variés (F_0, débit, intensité, coefficients M.F.C.C) a fourni des erreurs moins élevées (20%) pour les émotions simulées que pour la parole naturelle (M.Grimm, 2007).

Lorsque les émotions réelles de programmes audiovisuels sont notées selon les trois dimensions émotionnelles par des auditeurs et calculées par des estimateurs, on abouti aux erreurs les plus faibles lorsque ces derniers sont des vecteurs M.F.C.C par rapport aux estimations issues de paramètres d'énergie, de la fréquence fondamentale et de durée (D.Wu, 2010).

Les commentaires de courses hippiques ont fait l'objet d'analyses acoustiques montrant que les variations des paramètres sont similaires à celles de l'expression de la colère ou de l'allégresse (J.Trouvain, 2000).

Des enregistrements audiovisuels entre parents et enfants résolvant des problèmes personnels ont été annotés et classés pour être associés à des émotions. Deux caractéristiques acoustiques issues de la forme de l'onde glottique ont servi de descripteurs pour une reconnaissance automatiques des émotions. Il s'agit de l'aire de l'enveloppe spectrale du signal de parole et de celle de la forme de l'onde glottique pour des prononciations d'une seconde et demie environ. C'est pour la seconde que la reconnaissance aboutit à de meilleurs résultats avec des scores au-delà de 80%. Selon les auteurs, le résultat suggèrerait que la majeure partie du contenu émotionnel de la parole serait généré avant le processus de filtrage du conduit vocal (L.He, 2010).

Sources téléphoniques d'émotions réelles **:** L'enregistrement des conversations de deux personnes du centre de commande électrique de la ville de New York en charge de responsabilités importantes lors de la panne générale du 13 Juillet 1977 a fait l'objet d'analyses acoustiques (L.A.Streeter, 1983). Les résultats montrent que la fréquence

fondamentale du supérieur hiérarchique augmente tandis que celle de son interlocuteur diminue avec l'augmentation du stress.

Les centrales d'appels téléphoniques où des clients appellent un service spécialisé fournissent une matière émotionnelle réelle associée à l'irritation, la colère, l'angoisse mais aussi la satisfaction, l'excuse et la neutralité. Les analyses acoustiques ont montré que la gamme de variation de la fréquence fondamentale et la valeur maximale de la variation locale de F_0 entre deux segments voisés voisins interviennent fortement pour le classement des émotions rencontrées dans les dialogues (L.Devillers, 2003 et 19-24 Avril 2004). Les caractéristiques des pauses facilitent sensiblement mieux la distinction entre la colère et la peur que la fréquence fondamentale seule (L.Devillers, 23-26 Mars 2004). Plus récemment, les auteurs obtiennent de bons résultats pour la reconnaissance de la peur avec un mélange de descripteurs acoustiques et lexicaux (L.Devillers, 2006).

Préalablement à toute analyse, les messages retenus doivent être classés en catégories émotionnelles soit par des experts soit par un panel d'auditeurs. Vient ensuite l'étude de la relation entre les paramètres acoustiques et les labels émotionnels. C'est ainsi que neuf caractéristiques acoustiques ont été retenues comme pertinentes pour la détection de l'émotion en notant qu'elles sont plus nombreuses pour les voix féminines. Ces neuf paramètres sont : la durée maximale du silence dans les messages ; la valeur moyenne de l'énergie et sa gamme de variation ; celle de la distance euclidienne entre deux spectres successifs ; la valeur absolue du coefficient de tonalité (quotient entre les moyennes arithmétiques et géométriques du spectre à long terme) ; le centre de gravité spectral, la gamme de variation de F_0 et sa valeur maximale ; la durée en % des structures voisées (N.Chateau, 2004).

Pour un autre classement en émotions positives et négatives, des systèmes de reconnaissance automatique de cette valence ont abouti à des scores élevés, de l'ordre de 80% (C.M.Lee 2003 et 2005, L.Vidrascu 2005, D.Neiberg 2006). Pour la première étude, l'ajout de considérations lexicales aux descripteurs acoustiques diminue le taux d'erreur. Pour la seconde, les auteurs proposent vingt descripteurs pour la détection de la valence des émotions : il s'agit de caractéristiques liées à F_0 (pente, coefficient de régression, écart-type, gamme de variation etc), de caractéristiques des formants (fréquences moyennes des deux premiers, largeur de bande moyenne du premier, gamme de variation de la fréquence du second etc) et de paramètres énergétiques et de débit (L.Vidrascu, 2005). On peut également citer une labellisation des émotions réelles en positives, négatives, très négatives et neutres (C.Blouin, 2005).

Plus le nombre de classes émotionnelles augmente et plus les performances de la reconnaissance automatique des émotions ainsi labellisées décroissent (L.Vidrascu, 2007). La plupart des travaux ne retiennent qu'un faible nombre de classes, deux généralement en plus de la neutralité (positive et négative). La colère est parfois la seule classe labellisée et étudiée pour en effectuer la reconnaissance automatique (D.Morison 2007, D.Neiberg 2008, F.Burkhardt 2009).

Les descripteurs sont nombreux, plus d'une vingtaine, et variés (prosodiques, temporels et spectraux) et les taux de reconnaissance sont au-delà de 70%. On note qu'avec seulement cinq caractéristiques acoustiques (contour temporel de F_0 et de l'énergie, énergie du spectre en fonction du temps, largeur de bande des deux premiers formants) mesurées sur de courtes prononciations exemptes de développement linguistique, la qualité de la reconnaissance des émotions positives et négatives a dépassé également 70% (N.Baufadon, 2008).

Les probabilités de rencontrer des types variés d'émotion sont faibles dans les situations réelles décrites. L'objectif principal étant d'orienter automatiquement un appelant vers un opérateur humain si la détection émotionnelle en indique l'intérêt, la typologie réduite du classement répond à cette demande.

Récemment, il a été proposé d'abandonner la catégorisation des émotions en classes pour s'attacher à une estimation continue dans le temps de la valence et de l'activation des émotions grâce à des réseaux de neurones dotés d'une mémoire de l'historique du signal. Les résultats de la reconnaissance aboutissent à des taux d'erreurs de l'ordre de grandeur de ceux des auditeurs lors des tests perceptifs de reconnaissance (M.Wöllmer, 2008).

Ce résumé des travaux sur les émotions réelles, essentiellement issues de communications vers les centres d'appel ou d'émissions de télévision, peut être complété par les travaux d'analyse d'enregistrements d'appels téléphoniques d'urgence à la police (G.Demenko, 2008). Le propos est différent puisqu'il va s'agir de situations d'urgence avérée présentant une certaine similitude avec des situations aéronautiques captées par les enregistreurs phoniques.

Les résultats confirment la significativité des variations de la fréquence fondamentale pour identifier le stress malgré le nombre important de paramètres mesurés. Les auteurs remarquent que le décalage de F_0 est mieux corrélé avec le stress que sa gamme de variation. Les micro-tremblements des cordes vocales et les paramètres de rapport signal à bruit sont également sensibles. Il est à noter la difficulté à procéder à une catégorisation auditive des évènements en « dépression », « stress », « stress extrême ».

L'analyse acoustique des émotions réelles suit le même schéma d'étude que celui de l'analyse des émotions simulées. L'objectif étant d'en effectuer la reconnaissance automatique, leur classement s'impose. Mais dans ces cas là, la catégorisation devient plus floue, moins précise car les émotions ne sont pas aussi facilement identifiables, elles peuvent se superposer, les limites temporelles en sont variables et non contrôlées, et ceci même si l'intensité émotionnelle peut être forte.

Enfin, comment ne pas évoquer la question des choix de voix qui peuvent avoir lieu dans certaines professions liées aux centres d'appels téléphoniques mais également à celles de l'accueil, à celles du service au public et à celles de la direction d'entreprise.

« Souvenez-vous que l'intonation de la voix est très importante, parce que le ton, le débit et la clarté expriment votre attitude envers le consommateur. Vous ne devez jamais donner l'impression que vous vous ennuyez. Vos manières doivent donner l'impression que vous avez passé la journée à attendre cet appel en particulier. Pour vous y aider, essayez de sourire quand vous répondez à un appel » (A.Karpf, p. 475, 2008, extrait de « conseil aux employés des centres d'appel » cité dans D. Cameron, « Good to talk ? », Londres, Sage, 2000).

Ce « travail émotionnel » (A.Karpf, 2008), transforme la nature de la communication. La perte d'authenticité qu'est susceptible de déceler le demandeur peut influer sur l'expression de ses émotions. La notion de réalité des émotions peut perdre de son sens initial dans ces modes d'échange.

2.2.3. Stress Cognitif et Charge de Travail

Le stress cognitif désigne un stress reconnu directement par le système nerveux central (blessures physiques, émotions) différent du stress non cognitif ou physiologique perçu de façon secondaire (infection et inflammation) qui pourra s'exprimer sous forme psychique.

Les émotions en font donc partie mais les travaux des études répertoriées dans ce paragraphe réservent le terme à des situations expérimentales réelles ou provoquées où une difficulté est perçue, une pression temporelle s'exerce, des adaptations sont nécessaires. Il s'agit de la réalisation de tests cognitifs spécifiques, de tâches mentales, arithmétiques par exemple, plus ou moins complexes, accompagnées ou pas d'exercices psychomoteurs et du test de Stroop.

Le terme de charge de travail (« workload » ou « taskload ») fait référence à des expérimentations où un traitement croissant de la charge se produit pour le sujet, éventuellement pendant des tâches cognitives.

Stress Cognitif : Il a été provoqué par la mise dans l'obscurité totale et de légers chocs électriques accompagnant les réponses à des questions (L.Rubenstein, 1966). Pour une autre étude, il s'agissait d'effectuer une tâche arithmétique de sommation de nombres dans un temps donné et variable (M.H.L.Hecker, 1968). On note également la mise en place d'une tâche de localisation visuelle rendue difficile par des oscillations périodiques de l'éclairement de l'écran (J.D.Mosko, 1983). Des problèmes de logique à résoudre ont aussi été proposés (F.J.Tolkmitt, 1986) ainsi que des tests cognitifs spécifiques de construction de mots avec des lettres de l'alphabet, de répétition dans l'ordre et dans un ordre inversé de séquences de chiffres (P.Lieberman, 1995).

Il existe aussi le test de Stroop encore appelé test de stress. Après des phases classiques de lecture, le test propose d'énoncer la couleur avec laquelle est écrit un mot alors que celui-ci désigne lui-même une couleur différente. Le conflit cognitif qui en résulte entraine la hausse de la fréquence fondamentale et du jitter (L.J.M.Rothkrantz, 2004). Une récente étude montre une augmentation de la fréquence du premier formant et une décoissance de celle du second ainsi que des variations de la pente et de la durée des trajectoires temporelles de ces formants (T.F.Yap, 2010 et 2011). Les auteurs utilisent ces éléments comme descripteurs pour améliorer les performances d'une reconnaissance automatique du stress cognitif. Le test de Stroop a également servi de test pour un système de reconnaissance automatique basé sur le calcul par bande de fréquence de centroïdes spectraux et d'amplitude (P.N.Le, 2011).

Durant tous les types de tests, les sujets se sont exprimés et des paramètres acoustiques ont été mesurés. Sans entrer dans le détail des résultats spécifiques à chaque étude, il apparaît que le stress cognitif provoque des modifications de la voix.

Elles ont été mises en évidence par la comparaison détaillée de spectrogrammes montrant la diminution de la précision de l'articulation, des changements dans la quantité d'énergie aux hautes fréquences mais aussi par des mesures objectives d'accroissement de la fréquence fondamentale (M.H.L.Hecker, 1968), d'augmentation de la fréquence des deux premiers formants, de fluctuations plus importantes de l'amplitude des bruits de turbulence (friction et aspiration), de modifications de la durée des voyelles (J.D.Mosko, 1983). Des effets complémentaires du même type ont également été mesurés suivant le sexe du locuteur

et son degré d'anxiété naturelle (F.J.Tolkmitt, 1986). Enfin, le délai d'établissement du voisement (VOT) s'est révélé sensible et décroissant avec la montée en altitude de sujets conjointement avec la baisse de performance, lors de la réalisation de tests cognitifs en montagne (P.Lieberman, 1995).

Charge de travail : Ce qui à l'origine est une tâche, un exercice, devient une charge lorsque sa complexité s'accroit graduellement ou qu'elle se prolonge dans le temps. Les travaux qui ont été répertoriés traitent d'exercices psychomoteurs (M.Brenner 1987 et 1994, G.R.Griffin 1987), d'exercices mentaux arithmétiques (M.Alpert, 1988), de tâches de lecture (E.Mendoza, 1998 et 1999) ou de tâches cognitives spécifiquement élaborées (C.Baber, 1996).

Les analyses acoustiques vont s'attacher à étudier l'évolution de paramètres en fonction de la croissance de la charge de travail. Les résultats montrent que la fréquence fondamentale, l'amplitude maximale du signal suivent cette augmentation alors que la durée des mots diminue (le débit augmente) (G.R.Griffin 1987, M.Alpert 1988, M.Brenner 1987 et 1994). En revanche, ni le jitter, ni le schimmer ne répondent aux changements de charge de travail (M.Brenner, 1994). Ce n'est pas le cas pour une autre étude où le jitter et le shimmer augmentent tout comme la fréquence fondamentale et l'énergie à haute fréquence (E.Mendoza, 1998). Des paramètres de modulation d'amplitude et de fréquence du signal ont également montré des variations sous l'effet de la charge de travail cognitive (E.Mendoza, 1999).

La reconnaissance automatique a été expérimentée. Selon deux niveaux de rapidité de présentation, des sujets ont eu à lire des plaques d'immatriculation de véhicules avec l'alphabet O.A.C.I (Organisation de l'Aviation Cicile Internationale) (« Alpha, Bravo, Charlie … ») et les enregistrements ont été soumis à un système de reconnaissance automatique. Ses performances chutent avec l'augmentation du stress en même temps que la capacité des locuteurs à produire un discours acceptable par le système se détériore (C.Baber, 1996). Ces mêmes auteurs ont constaté le même résultat avec une tâche d'identification et de classement de formes complétée de calcul arithmétique avec deux niveaux de charge de travail. Il a été relevé une augmentation du débit phonatoire avec le niveau de charge.

L'amplitude des micro-tremblements des cordes vocales a été modifiée par une charge de travail cognitive issue d'une tâche de lecture rapide avec un retard audiophonatoire (E.Mendoza, 1999).

Enfin, une étude plus récente compare la charge cognitive (tâches arithmétiques) au stress émotionnel (à la vue de photographies) et semble indiquer que le débit, la fréquence fondamentale moyenne, l'énergie à haute fréquence (500 à 1600 Hz), les gradients de croissance et de décroissance de l'énergie sur l'enveloppe du signal augmentent, alors que pour le stress psychologique seule la fréquence fondamentale et la distribution spectrale de l'énergie varient (K.R.Scherer, 2002).

Le nombre d'études répertoriées n'est pas très important, bien qu'il en existe certainement d'autres, mais il est suffisant pour conclure à l'existence des modifications vocales par l'effet d'un stress cognitif et d'une charge de travail.

2.3. Voix et Fatigue (références bibliographiques § 5.2.3. page 139)

Sous l'appellation fatigue, sont regroupés ici les travaux qui en traitent en tant que telle mais aussi ceux qui concernent la vigilance (« alertness »), la somnolence (« drowsiness »), conséquences de la fatigue. Comme il a été vu précédemment (§ 2.1 page 8), plusieurs études acoustiques dans le domaine des transports s'inscrivent dans ce champ (J.Whitmore, 1996 ; P.Shahidi, 2010 ; K.Shiomi 2000, 2008 et 2009).

Celles qui suivent traitent de la fatigue provoquée par un effort physique comme la course sur un tapis roulant (H.S.Hayre, 1980), un escalier automatique (S.A.Patil, 2008), mais aussi la privation de sommeil (Y.Harrison 1997, T.Nwe 2006 ; H.P.Greeley, 2006 et 2007 ; J.Krajewski 2009, A.D.Bagnall, 2007 et 2011). Elles consistent à examiner le signal vocal. Il existe une autre voie de recherche fondée sur la mesure de caractéristiques visuelles du visage (Q.Ji, 2004). En 1932, la fatigue a été examinée par une mesure de l'augmentation du tremblement musculaire suite à l'augmentation d'un poids à soulever (W.A.Bousfield). Cette étude est mentionnée car elle est à l'origine des travaux sur les micro-tremblements de la voix pour la détection du mensonge (cf. § 2.4.1 page 43).

Les résultats des analyses acoustiques montrent une diminution de la fréquence fondamentale de quelques Hertz (A.D.Bagnall, 2007 et 2011), et des modifications des valeurs numériques de coefficients cepstraux (M.F.C.C) dont la corrélation avec l'état de repos chute avec la durée d'absence de sommeil (H.P.Greeley, 2007). Une classification reposant sur un vecteur des amplitudes issues d'un banc de filtre a permis d'isoler les prononciations en l'absence de sommeil (T.Nwe, 2006). Avec un grand nombre de descripteurs acoustiques présentés sous forme de contours temporels, une classification automatique de la voix en l'absence de sommeil a permis d'atteindre des scores de reconnaissance de plus de 85% (J.Krajewski 2009). La voix devient plus monotone et les individus n'utilisent pas les intonations appropriées (Y.Harrison 1997). Des tests d'écoute révèlent que des auditeurs perçoivent le caractère fatigué de la voix lorsqu'une privation de sommeil a eu lieu (A.D.Bagnall, 2011).

La fréquence des quatre premiers formants de plusieurs voyelles (H.P.Greeley, 2006) s'est révélée elle aussi posséder quelques fortes corrélations avec la durée de la privation de sommeil mais pas pour toutes les voyelles et tous les formants (H.P.Greeley, 2006).

Les auteurs ont également montré la baisse du degré de confiance d'un système de reconnaissance automatique de la parole en situation de fatigue du locuteur.

La reconnaissance automatique de la fatigue physique (et du stress) grâce à un algorithme de calcul de l'opérateur de Teager (« Teager Energy Operator ») a montré des taux de reconnaissance de l'ordre de 70% (S.A.Patil, 2008).

Enfin, une mesure de deux paramètres spectraux corrélés au rythme cardiaque et indiqués comme aptes à détecter la fatigue n'est pas exploitable car l'auteur n'en donne pas la définition en précisant qu'ils sont sa propriété (H.S.Hayre, 1979). Des modifications de la structure formantique ont aussi été relevées sous l'effet de la fatigue (et de l'émotion) sans que soient précisées les conditions expérimentales ni la nature des perturbations (J.Davidović, 1997).

Plus récemment, des travaux sur le caractère chaotique de la voix ont eu pour objet de détecter la fatigue d'un locuteur (K.Shiomi, 2003, 2004, 2005, 2008, 2009). Pour ce faire, les auteurs proposent et commercialisent un appareil spécifique, le « CENTE - cerebral resource and activity measurement equipment – fatigue and drowsiness predictor ».

Il est destiné à évaluer le niveau de fatigue non seulement des conducteurs automobiles (individus et taxis), de camions, de transport public (pilotes d'avion et contrôleurs aériens, conducteurs de trains, autocars) mais aussi des médecins et des individus menant des activités intellectuelles intenses. Il peut être également utilisé dans les cas d'accumulation de stress.

Son application principale concerne la fatigue mais, n'étant pas destiné exclusivement au secteur des transports, les travaux de recherche qui lui ont donné naissance sont examinés dans ce paragraphe plus général « voix et fatigue ».

C'est dans l'observation dans le plan des phases de voyelles prononcées en état de fatigue que débutent les travaux (K.Shiomi, 2000). Leurs différences sont numériquement mises en évidence par le premier exposant de Lyapunov λ qui augmente graduellement pendant les 30 minutes de l'expression orale d'un locuteur se déclarant fatigué par l'exercice vocal de lecture à haute voix. Cette augmentation est confirmée par des expériences du même type sur de longues durées de prononciation avec trois locuteurs (K.Shiomi, 2004). Un premier brevet a été déposé pour signifier la réalisation d'un appareillage de mesure de la fatigue et de la somnolence fondé sur la comparaison de λ à différents instants (K.Shiomi, 2005).

Les auteurs définissent ensuite un ensemble de paramètres appelés « cerebral exponents » et mesurés par un algorithme spécifique « Shiomi's cerebral exponents calculation algorithm » (SiCECA) pour évaluer l'activité cérébrale du locuteur (K.Shiomi, 2003 et 2004). Ils sont issus du constat que la dynamique du système de production de la parole n'est pas stable même pendant la prononciation d'une voyelle alors que la stabilité est un préalable au calcul de λ par les méthodes classiques (H.Kantz 2005 dans §5.2.5 page 142). Il faut noter que les auteurs traitent les voyelles dans leur intégralité, attaque et extinction comprises, et ne prennent pas en considération la seule tenue orale quasi-stationnaire. À cela s'ajoutent les consonnes du langage puisqu'ils analysent en continu l'intégralité de la parole. L'ensemble possède en effet une dynamique instable (K.Shiomi, 2008).

La différence principale avec l'estimation de λ réside dans le choix de l'intervalle de temps pour le calcul du voisinage de chaque point de la trajectoire. Avec les méthodes classiques ce choix est fixé préalablement aux calculs alors qu'ici il est variable d'un segment analysé à l'autre. Le premier échantillon appartenant au voisinage est situé à presque une période du signal et ainsi de suite. Cet intervalle de temps pour le recensement du voisinage est compris entre la plus petite et la plus grande période du signal.

Les résultats d'expériences avec des sujets effectuant de la simulation de conduite de train et des exercices physique pendant cinq jours consécutifs montrent des variations de l'exposant cérébral en relation avec la fatigue (K.Shiomi, 2008 et 2009).

La fatigue vocale a été examinée en demandant à des acteurs de lire trois heures de texte et a montré que les nasales (voyelles et consonnes) paraissaient être les plus sensibles aux effets sur la base de mesures de paramètres prosodiques et spectraux (M.J.Caraty, 2010) et d'une analyse discriminante.

Même si le nombre de travaux est en nombre restreint, il ne fait pas de doute que la parole est modifiée par la fatigue. Les récentes recherches de l'équipe de K.Shiomi sont très séduisantes par l'affirmation des auteurs à évaluer l'état du locuteur. Mais des résultats expérimentaux font défaut notamment sur l'analyse de cas réels et pour tous les divers types de métiers pour lesquels ils préconisent l'utilisation de leur système de détection.

2.4. Voix et Autres Facteurs Perturbateurs (références bibliographiques § 5.2.4. page 141)

On distingue la détection du mensonge, les pathologies qui ne sont pas en relation avec l'appareil phonatoire et la détection de l'état d'ébriété.

Sauf pour la détection du mensonge, la revue des connaissances est ici beaucoup moins riche que pour les paragraphes précédents et les publications sont moins abondantes. Le propos est de citer les autres types de désordres psychophysiologiques que la voix peut mettre en évidence en les illustrant de quelques exemples.

Hors de ce classement, une étude s'est attachée à examiner le caractère charismatique du discours de certains hommes politiques (A.Rosenberg, 2005). La procédure a consisté à corréler des scores issus des impressions subjectives d'auditeurs à la fréquence fondamentale, le niveau sonore et le débit phonatoire des orateurs. La corrélation s'est avérée forte pour chacun des trois paramètres lorsqu'ils possèdent des valeurs élevées.

2.4.1. Voix et Mensonge

L'intérêt de la détection vocale du mensonge repose sur le caractère non invasif de la prise de son contrairement à des mesures plus classiques que sont les activités cardiovasculaire (fréquence cardiaque, pression sanguine et volume sanguin), respiratoire et électrodermale mesurées par le polygraphe.

Les premières études acoustiques remontent probablement à 1963 et 1964 où l'énergie du signal de parole respectivement dans deux et trois bandes de fréquences est mesurée (M.Alpert, 1963 ; A.J.Friedhoff, 1964). Les auteurs ont remarqué des changements de la distribution de l'énergie dans le spectre lors du mensonge.

D'autres recherches ont donné lieu à la commercialisation de plusieurs appareillages réalisant la mesure de micro-tremblements des cordes vocales (A.D.Bell, 1976 ; H.Charles, 2005 ; P.O'Neal Nunally, 2007, J.D.Williamson 1979) ou sur celles d'autres paramètres :

- le nombre et la largeur des « plateaux » et des « épines » sur l'allure temporelle du signal qui permettent de statuer sur le mensonge ou pas mais aussi sur certains traits émotionnels (évasif, sarcastique …) (A.Liberman, 2003) ;
- la comparaison des amplitudes maximales du signal temporel dans deux bandes de fréquence (150-300 Hz et 600-1200 Hz) (F.H.Fuller, 1974).

La théorie liée aux micro-tremblements repose sur le fait qu'une oscillation d'environ 10 Hz accompagne la contraction des muscles (O.Lippold, 1971) après un effort physique (O.Lippold, 1976). En 1971, un appareillage a été conçu par la firme « Dektor Counterintelligence and Security », le « Psychological Stress Evaluator » (P.S.E), signifiant

qu'il mesurait ces tremblements physiologiques dans la voix et qu'ils étaient atténués par l'effet du stress psychologique (A.D.Bell, 1976). Cet appareil est devenu depuis le « Computer Voice Stress Analyzer » (CVSA) distribué par le « National Institute of Truth Verification ».

Après ralentissement de la bande magnétique de l'enregistrement, la mesure réalise une démodulation en fréquence du signal de parole en transformant la modulation de fréquence (entre 8 et 12Hz) en une modulation d'amplitude (circuit résonnant accordé). Si la fréquence augmente, le niveau de sortie augmente. Après une détection d'enveloppe, le système fournit un diagramme qui possède naturellement des ondulations irrégulières devenant très régulières en état de stress ou de mensonge. Les variations de fréquence présentes à l'état de repos disparaissent lorsque le sujet est troublé.

Des versions numériques de la mesure ont été introduites pour affecter un score aux diagrammes obtenus afin de s'affranchir de l'interprétation strictement visuelle des schémas (H.Charles, 2005). Le calcul a été également programmé sur Matlab® (J.H.L.Hansen, 1999).

Les tests ont été nombreux et les résultats controversés. Les fabricants attestent de la véracité de la mesure, une étude relève une bonne corrélation des diagrammes P.S.E avec les performances de sujets réalisant une tâche psychomotrice (S.G.Schiflett, 1980) mais d'autres études ne lui attribuent pas la même fiabilité que ce soit pour la détection du mensonge (M.Brenner, 1979, V.L.Cestaro, 2002, F.Horvath, 1978) ou celle de l'anxiété (G.A.Smith, 1977). Un travail conclue de manière plus radicale en ne reconnaissant pas la qualité de l'appareillage pour la détection du mensonge, inapte à différencier ce dernier du stress émotionnel (D.Haddad, 2002).

Une mesure similaire a été proposée, non pas pour rechercher une modulation de fréquence, mais une modulation d'amplitude du signal vocal (F.H.Fuller, 1974 ; W.S.Winholtz, 1992). Le premier brevet s'attache à détecter le mensonge et la seconde étude est appliquée au cas de pathologies non vocales (maladie de Parkinson). Cette dernière s'attache également à mesurer une modulation de fréquence. Les résultats montrent des variations significatives entre les groupes de sujets testés pour la fréquence de la modulation d'amplitude mais pas pour celle de la modulation de fréquence.

Enfin, la fréquence fondamentale moyenne s'est révélée être plus élevée en répondant volontairement mensongèrement à des questions qu'en prononçant la vérité (5 Hz environ) (L.A.Streeter, 1977). L'usage des pauses dans la conversation est moins corrélé avec le mensonge qu'avec l'expression de la vérité probablement du à un plus grand contrôle du sujet lorsqu'il décide de mentir (S.Benus, 2006).

2.4.2. Voix et Pathologies Non Vocales

Il s'agit de pathologies psychologiques comme la dépression (A.Nilsonne, 1988) pour laquelle les auteurs ont trouvé une corrélation entre l'état clinique et la fréquence fondamentale F_0 (ou plus exactement avec le taux moyen de variation de F_0 pendant des intervalles de 10 ms et avec l'écart-type de cette mesure). Comme indiqué dans le paragraphe précédent, la mesure d'une modulation d'amplitude a été appliquée pour la maladie de Parkinson (W.S.Winholtz, 1992). Pour cette même maladie des mesures acoustiques ont

révélé une augmentation de la fréquence fondamentale et une diminution de sa variabilité (J.K.Y.Ma, 2010)

La scizophrénie a fait également l'objet d'étude sur la voix (R.Cowie, 1996).

Deux autres études montrent les relations entre ces désordres mentaux et l'expression vocale même si des mesures acoustiques n'ont pas été réalisées et qu'il s'agit de l'estimation auditive de la qualité de la voix selon des caractéristiques subjectives prédéterminées (C.Chevrie-Muller, 1978 et 1985).

Les variations à la hausse de la fréquence fondamentale et du débit ont été également mises en relation avec des états psychologiques persistants de colère et de tristesse (D.Rochman, 2008).

Les pathologies auditives ont également donné lieu à des mesures acoustiques de la voix des déficients auditifs (R.Cowie 1996, E.Scheimer 2005).

On notera enfin les articles de l'ouvrage « Speech evaluation in psychiatry » (J.K.Darby, 1981) qui décrivent les liens entre la voix et les maladies psychiatriques démontrant que cette dernière est modifiée par ces pathologies.

2.4.3. Voix et Alcoolémie

Deux études ont été menées sur le sujet en relation avec le secteur des transports puisqu'il s'agissait d'étudier la voix du capitaine du pétrolier accidenté Exxon Valdez.

Une expérimentation avec des sujets volontaires dont le taux d'alcool dans le sang était mesuré en laboratoire a permis de mettre en évidence des modifications vocales (F.Klingholz, 1988). Les paramètres mesurés étaient la distribution fréquentielle de la fréquence fondamentale F_0, le rapport signal sur bruit de voyelles, le spectre à long terme et le rapport de la fréquence des deux premiers formants (F_1/F_2).

Grâce à une méthode de classification utilisant le calcul de distances euclidiennes entre les distributions de valeurs des paramètres en état d'ébriété et celles en état normal, les auteurs ont pu montrer que F_0, le rapport signal sur bruit et le spectre à long terme pouvaient discriminer les deux états.

Les variations de mesures liées au débit et à la fréquence des trois premiers formants se sont révélées significativement différentes pour les prononciations alcoolisées et une tendance à l'hyper articulation a été relevée (F.Schiel, 2010). La durée des phrases et l'écart-type de la fréquence fondamentale augmentent suggérant pour ce dernier point un contrôle moins précis de la vibration des cordes vocales (D.B.Pisoni, 1989).

D'un point de vue perceptif, les auditeurs sont capables d'identifier l'existence et l'augmentation de l'état d'ébriété sans pour autant être à même d'en estimer des niveaux spécifiques (H.Hollien, 2009). Ils peuvent différencier les prononciations en état d'ébriété de celles sans prise d'alcool lorsqu'elles sont présentées par paire ou seules (D.B.Pisoni, 1989).

2.5. Synthèse de la Revue Générale des Connaissances

L'ensemble des travaux montre que l'origine des perturbations est variée. La diversité s'exprime par la multiplicité des émotions susceptibles de modifier la parole, qu'il s'agisse d'émotions simulées ou réelles. Cette diversité est également reflétée par l'existence de nombreuses autres causes possibles, comme la fatigue ou le stress cognitif, et la mise en œuvre de multiples expérimentations.

Les études démontrent que le signal vocal se modifie sous l'effet de ces divers facteurs perturbateurs. Les mesures en attestent et confirment les impressions subjectives de l'écoute.

Ces deux points généraux sont fondamentaux car ils ne mettent pas en doute l'existence du contenu non linguistique du signal vocal. Mais la relation entre les causes et les effets acoustiques n'est pas biunivoque. Lorsque des modifications sont mesurées, il n'est pas encore possible d'en identifier la cause générale (type d'émotion, fatigue …) et d'y associer un degré.

En d'autres termes, l'analyse d'un signal ne permet pas, à elle seule, de statuer sur l'existence et l'origine des perturbations psychophysiologiques qui auraient donné naissance aux supposées modifications.

Pour l'essentiel, ce n'est qu'en disposant d'extraits de voix neutres, ou au repos, qu'il est envisageable d'identifier des altérations du signal. Les implications sont fortes pour de futures détections en temps quasi réel pour lesquelles il serait souhaitable de disposer de ces éléments de parole. L'écoute fournit également des éléments d'appréciation importants, qu'il s'agisse de celle de l'expérimentateur ou de celle de jurys d'auditeurs. Le système auditif en possède l'aptitude et il peut également se servir des éléments linguistiques pour appuyer son jugement.

En matière de reconnaissance automatique des émotions, la problématique est similaire. Sans références, la tâche se complique et les performances chutent. Cette reconnaissance a fait ses preuves dans le cas d'émotions simulées par des acteurs, beaucoup moins dans des cas réels surtout si les perturbations ne peuvent pas être affectées à des catégories précises.

L'analyse acoustique reste le socle des méthodes de reconnaissance. Il semble encore prématuré de considérer que chaque type de perturbation psychophysiologique puisse être associée à un vecteur acoustique type possédant des valeurs de paramètres bien déterminées, avec des tolérances.

Malgré le très grand nombre de caractéristiques testées dans des situations expérimentales différentes, il semble que la taille du vecteur, facilitée par l'accroissement de la puissance de calcul, réponde au souci de fournir au système de reconnaissance le plus de données possibles pour lui faciliter la décision.

Le problème se trouve déplacé dans les champs informatique et mathématique sans que la résolution de la problématique acoustique soit achevée. Par exemple, on ne dispose pas de valeurs seuils, de domaine de variation ou de critères propres à chaque type de perturbation, d'ailleurs susceptibles d'améliorer les performances de la reconnaissance automatique.

La trace vocale de ces perturbations ne possède pas son équivalent en termes de caractéristiques du signal. On peut s'interroger sur son existence auditive. L'auditeur repère très certainement les modifications sans pour autant, sauf si on le lui demande, procéder à une catégorisation, une reconnaissance du trouble en le nommant en tant que tel.

La stratégie des recherches a donc fait la place à la reconnaissance automatique sans que les phénomènes acoustiques soient bien maitrisés. Ils ont été identifiés mais n'ont pas conduit à l'élaboration de modèles.

Il n'existe pas de modèle de l'émotion ou de la fatigue en fonction de paramètres ou de combinaisons de caractéristiques acoustiques comme il peut exister des modèles de l'intelligibilité. Il est vrai que cette dernière peut être plus facilement quantifiée par des tests de reconnaissance auditive des unités du langage. Mais rien n'interdit de penser que des mesures acoustiques avec des tests subjectifs d'écoute et des mesures physiologiques ou psychologiques, ne puissent pas être en mesure de fournir la base nécessaire à l'élaboration de modèles.

Hormis dans le cas des émotions simulées, la littérature est absente de compte-rendus d'expérimentations à grande échelle. Dans le domaine des transports, directement concerné par ces recherches, ce type d'étude n'existe pas.

En revanche, l'ensemble des travaux porte sur un très grand nombre de sujets, y compris de manière interculturelle, et s'accompagne de la création de nombreuses bases de données. C'est d'ailleurs ce qui permet d'affirmer les deux points fondamentaux de synthèse du début du paragraphe.

L'examen des travaux suscite l'impression que la recherche a surtout privilégié l'étude des émotions reproduites ou réelles, au sens où le psychologue les identifie, pour des besoins de reconnaissance ou de synthèse sonore sans mettre l'accent sur de plus subtiles et fines perturbations de l'état de neutralité.

Ils ont été nécessaires pour comprendre comment le signal est modifié, quelles sont les familles de caractéristiques sensibles et dans quelle mesure. Mais les résultats ne sont pas directement transposables pour effectuer une détection dans des applications comme celles qui concernent le domaine des transports où la perturbation émotionnelle est beaucoup moins bien identifiée voire localisable.

3. CONTRIBUTION de l'ANALYSE ACOUSTIQUE de la VOIX au CONTEXTE AÉRONAUTIQUE et au DOMAINE de la PSYCHOLOGIE

Comme précisé dans le paragraphe introductif définissant le cadre de la recherche, la contribution à la thématique est double.

- Les travaux réalisés complètent l'ensemble des études en la matière en particulier dans le contexte aéronautique où les travaux sont relativement peu nombreux.

- Trois caractéristiques originales sont proposées et une modélisation est introduite.

Les expérimentations mises en œuvre, les paramètres mesurés et les résultats reprennent la classification établie pour la synthèse des connaissances: « Émotion » (§ 3.1 page 49) et « Fatigue » (§ 3.2 page 62). Des expériences complémentaires relevant du domaine de la Psychologie sont ensuite évoquées (§ 3.3 page 74).

Les deux premières parties ciblent les deux types de perturbations possibles de l'état général des pilotes d'avion. Les divers contextes expérimentaux étudiés en spécifient et en précisent la nature. Il ne s'agit pas d'émotion ou de fatigue au sens le plus large mais de leur déclinaison dans le milieu aéronautique.

Les deux expérimentations qui sont indiquées comme relevant du domaine de la Psychologie se situent en dehors de ce contexte.

La première (test de Stroop) entre dans le cadre d'une collaboration académique. Pour la seconde, il s'est agit d'évaluer la pertinence d'un champ de recherche original dans le domaine des modifications vocales, celui des relations possibles entre la vision et la voix.

Ces deux expériences ont une dimension psychologique parce qu'elles sont respectivement fondées sur un test mis au point par les psychologues et par l'attribution de notes émotionnelles à des images par les spécialistes de la discipline.

Pour tous les travaux mentionnés dans cette contribution, le détail des conditions expérimentales, des conditions d'analyse et des résultats sont mentionnés dans les publications correspondantes et ne sont pas intégralement décrits ici. Le propos est d'en récapituler les grandes lignes en mettant l'accent sur les résultats les plus significatifs.

La synthèse de l'ensemble de ces travaux expérimentaux fait l'objet du paragraphe 3.4.

L'avant dernier paragraphe (§ 3.5 page 81) présente les méthodes et les conditions de mesure des paramètres acoustiques classiques dans les situations expérimentales mises en œuvre puis développe les caractéristiques originales et la modélisation. L'exposé de la contribution se conclut par une synthèse critique des travaux théoriques (§ 3.6 page 112).

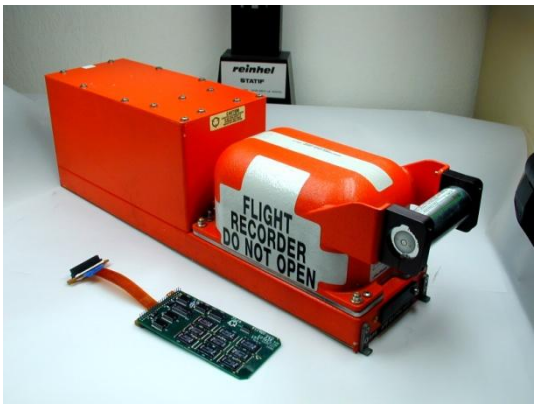
3.1. Émotion

L'apparition de pannes à bord d'un aéronef peut induire du stress chez le pilote. Ce stress ou cette émotion intéressent l'étude lorsqu'ils suscitent un dialogue dans le poste de pilotage. En l'absence de suffisamment d'éléments de vocabulaire, les analyses sont plus difficiles car la représentativité du suivi temporel et des moyennes de paramètres par phase dépendent du nombre d'échantillons de parole. De ce point de vue, le choix particulier d'un enregistrement en vol des conversations lors d'un accident aéronautique a fournit une matière vocale dense avec une répétition d'incidents.

En conditions contrôlées, l'utilisation d'un simulateur de vol permet de provoquer à souhait des incidents de vol et de garantir l'existence d'une quantité suffisante d'échanges verbaux.

3.1.1. Enregistreur phonique des conversations en cockpit d'avion accidenté (partenaire : Bureau d'Enquêtes et d'Analyses B.E.A), 1989 à 1992, [8], [9], [10] et 2010 [35]; Une partie de l'étude a fait l'objet du co-encadrement d'un étudiant en D.E.A (P.Vidal 1994) ;

Un enregistrement a été sélectionné car les dysfonctionnements constatés avant l'accident ont provoqué de longues discussions entre les membres d'équipage offrant ainsi une matière sonore riche pour l'étude des conséquences vocales de l'émotion.



Enregistreur de vol embarqué et sa mémoire électronique (source : Bureau d'Enquêtes et d'Analyses)

Les enregistreurs de vol embarqués sont deux équipements permettant d'enregistrer les informations sur le vol. L'enregistreur de paramètres enregistre divers paramètres de vol par rapport à une base de temps définie et l'enregistreur phonique enregistre l'environnement acoustique du poste de pilotage (conversations, annonces, échanges radio-téléphoniques, alarmes et bruits divers). Le C.V.R (Cockpit Voice Recorder ou enregistreur phonique) enregistre en boucle fermée sur trente ou 120 minutes les échanges entre les pilotes ou avec les contrôleurs, les bruits et les alarmes sonores. Selon leur génération, les enregistreurs utilisent deux types de support : la bande magnétique quart de pouce ou la mémoire statique (http://www.bea.aero).

Quatre pistes d'enregistrement existent :, celle de la base de temps incluant le « public address » (communications entre la cabine passagers et la cabine de pilotage) et le cas échéant le microphone d'un troisième membre d'équipage, celle du microphone à bouche du

Commandant de bord, celle du microphone à bouche de l'Officier pilote de ligne (copilote) et celle du microphone d'ambiance placé au milieu du poste de pilotage.

Les spécialistes acousticiens et les enquêteurs du Bureau d'Enquêtes et d'Analyses (B.E.A) transcrivent les enregistrements phoniques et effectuent des analyses spectrales qui participent à l'identification des phénomènes sonores. Lorsque l'enregistreur des paramètres de vol ne fournit pas d'informations complètes et fiables parce que certains paramètres sont temporairement mal enregistrés ou que leur fréquence d'échantillonnage ne permet pas une étude détaillée de leur comportement, l'enregistreur phonique peut être d'une aide précieuse pour recalculer par exemple les vitesses de rotation des moteurs à partir de l'enregistrement sonore ou pour identifier les alarmes et des bruits comme ceux de sélecteurs mécaniques dans le poste de pilotage.

Mais aucune analyse vocale n'est encore effectuée. La faisabilité de ces analyses est confiée à la recherche universitaire qui dans le cas présent ne s'enquiert pas des sons d'ambiance.

La qualité sonore des enregistrements est généralement propice aux analyses acoustiques de la voix des personnels navigants. La protection des enregistreurs et les précautions mises en œuvre pour leur dépouillement sont telles qu'ils livrent correctement les informations stockées pour la très grande majorité des accidents. La qualité intrinsèque des enregistreurs phoniques à bande magnétique comprend une bande passante de 150 – 5000 Hz. Elle est de 8 kHz pour la piste du microphone d'ambiance (jugée la plus importante) des enregistreurs à mémoire statique, et imposée par une fréquence d'échantillonnage de 16 kHz. Les pistes des microphones à bouche des membres d'équipage sont échantillonnées à 8 kHz. Ils sont équipés d'un contrôle automatique de gain et sont de fait moins sensibles aux niveaux sonores des bruits ambiants. Leur position près de la bouche du locuteur renforce cette propriété.

Outre l'examen des variations chronologiques du niveau sonore pour cause de variation automatique du gain dans la chaîne électroacoustique, ces considérations rendent possibles les analyses du signal vocal.

La qualité acoustique des microphones à bouche est quant à elle très variable puisque d'une compagnie à l'autre, les équipements sont différents. De plus, les transducteurs ne sont pas choisis pour répondre à des besoins d'analyse du signal vocal mais pour garantir une intelligibilité correcte de la parole. Les exigences acoustiques sont moindres. Par ailleurs, il peut arriver que les pilotes ne soient pas équipés de leur dispositif microphonique avant l'accident. Dans ce cas, c'est le microphone d'ambiance qui fournit le détail des conversations.

En dehors de ces questions de qualité sonore des enregistrements, toutes les situations d'accident ne se prêtent pas à une étude vocale des facteurs humains susceptibles d'en être une cause. La situation aéronautique en tant que telle, conditionne l'intérêt d'une analyse des voix. Outre la période critique de fin du vol, sont propices les contextes de vol où des discussions existent et où il existe suffisamment d'éléments de parole. Ce n'est pas lorsque l'évidence de l'issue du vol se présente qu'il est nécessaire de s'enquérir des modifications vocales. Sans en occulter l'étude, la priorité est plutôt donnée aux phases de vol qui précèdent pour leur lien avec les facteurs humains susceptibles d'être la cause totale ou partielle de l'accident. L'intérêt de l'étude est également d'envisager la corrélation de la détection des

causes humaines des accidents avec les informations fournies par l'enregistreur des paramètres de vol.

L'enregistrement étudié ici a constitué la situation expérimentale de la thèse de doctorat. Depuis, les mesures ont progressé puisque le nombre de paramètres acoustiques a augmenté et qu'ils sont désormais estimés à partir de programmes informatiques de laboratoire plutôt qu'avec un logiciel dédié. Il a donc à nouveau été analysé [35]. Cet enregistrement présente un grand intérêt car il traite d'une situation réelle de stress pendant plusieurs minutes et qu'il offre également quelques minutes de prononciations en situation normale.

La chronologie de l'enregistrement est la suivante :

- 1. Phase d'atterrissage (9 minutes) : les pilotes sont en approche de l'aéroport et se posent. Les conditions de vol sont apparemment délicates compte tenu de la météo. C'est ce que l'on comprend à partir de leurs échanges après que l'avion se soit posé. Un dysfonctionnement dans les commandes de vol est repéré à l'atterrissage. Un objet tombe dans la cabine après l'atterrissage.

- 2. Phase de décollage : les conversations enregistrées reprennent lorsque le même équipage se prépare pour le décollage. On distingue les périodes suivantes :

  2.1 mise en route des moteurs, procédures de décollage (« checklist ») et début du roulage (9 minutes) ;

  2.2 de nouveau apparition du problème de commande ; les pilotes en discutent, décident de faire un nouvel essai sur la bretelle et poursuivent les procédures à réaliser (2 minutes) ;

  2.3 nouvelle apparition du dysfonctionnement à la mise en puissance pour le décollage ; décollage interrompu.

  2.4 discussion entre les membres d'équipage qui les conduit à évoquer l'éventualité d'une origine aérodynamique au problème, à aller observer à l'arrière de l'avion par un hublot s'il n'y a rien d'anormal. Ils procèdent à nouveau à la « checklist » avant décollage car le problème ne se manifeste plus (10 minutes) ;

  2.5 décollage et accident fatal (trois morts) après l'envol de l'appareil (1 minute).

L'intérêt de la situation aéronautique tient à la répétition du dysfonctionnement et aux longues discussions entre les trois membres de l'équipage (Commandant de Bord CDB, Officier Pilote de Ligne OPL et Ingénieur navigant ING).

Le résultat le plus significatif concerne la fréquence fondamentale F_0 dont les variations à la hausse peuvent être rapprochées du contexte émotionnel.

Avec 1236 voyelles segmentées dont 602 pour le CDB, 405 pour l'OPL et 229 pour l'ING, une analyse minutieuse des variations a pu être mise en correspondance avec le sens des phrases et la situation dans le poste de pilotage.

À titre d'exemple, la figure 3.1.1.1 illustre ce que sont les variations temporelles de F_0 pour l'OPL. Chaque point représente la fréquence fondamentale moyenne d'une voyelle segmentée.

Les fluctuations de F_0 tout au long de la bande se situent entre 80 et 300 Hz environ. La valeur moyenne pour les 405 voyelles segmentées est de 143,5 Hz (σ = 40,2 Hz).

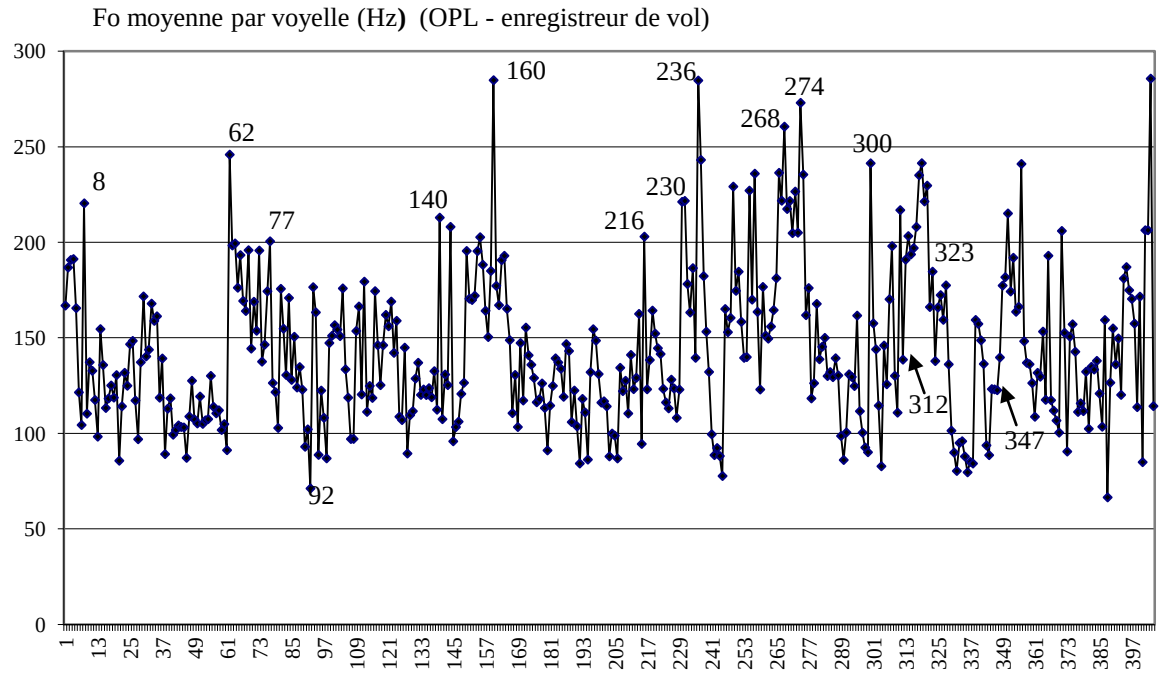


Figure 3.1.1.1 : Fréquence fondamentale moyenne de chaque voyelle prononcée par l'officier pilote de ligne (le n° d'ordre de quelques voyelles est indiqué) [35].
Voyelles 1 à 83 : phase 1 d'atterrissage ;
Voyelles 84 à 229 : préparation du décollage (phases 2.1, 2.2 et 2.3) ;
Voyelles 230 à 275 : discussion entre les membres d'équipage (phase 2.4) ;
Voyelles 276 à 312 : décollage et accident fatal (phase 2.5).

La phase d'atterrissage de l'appareil s'étend jusqu'à la voyelle 83. Il s'y détache une zone de valeurs élevées qui correspond aux voyelles 62 à 77. Elles sont prononcées lorsque l'avion est au sol et l'OPL parle des difficultés rencontrées pour atterrir. Il n'est donc pas possible d'attribuer ces valeurs élevées à un stress. Il peut s'agir d'une manifestation émotionnelle bien postérieure à l'apparition du problème car le pilote cherche à justifier son point de vue sur la question mais la corrélation avec le contexte est délicate. La F_0 moyenne de la phase d'atterrissage est égale à 137,3 Hz (σ = 34,7 Hz) et elle devient égale à 127,3 Hz (σ = 28,3 Hz) si on ne prend pas en compte les voyelles 62 à 77. On trouve ici une illustration de l'effet gommant des moyennages par groupes de voyelles.

Pour le CDB aucun stress particulier n'est à relever pendant la phase d'atterrissage. Il n'existe pas de F_0 élevées dans ces propos. Il est vrai qu'il n'avait pas la charge d'effectuer les principales opérations pour poser l'appareil.

À partir de la voyelle 84 commence la phase de préparation au décollage. Il apparaît que jusqu'à la voyelle 140 les variations de F_0 sont de la même nature que celles des voyelles 1 à 61 de la phase d'atterrissage. La valeur moyenne est d'ailleurs égale à 131,6 Hz (σ = 26,3 Hz).

Deux voyelles isolées 140 et 144 possèdent des valeurs de F_0 supérieures à 200 Hz sans que le contexte puisse particulièrement en être la cause, il s'agit de variations probablement naturelles. En revanche, l'ensemble des voyelles 148 à 160 constitue une zone de fréquence fondamentale élevée en lien direct avec le contexte puisqu'il s'agit de la première phrase que prononce l'OPL lorsqu'il constate le problème de commande. La F_0 moyenne pour les 13 voyelles concernées est de 178,9 Hz (σ = 40,7 Hz). En éliminant la voyelle 160 du calcul de la moyenne car elle possède une F_0 particulièrement élevée, la moyenne reste proche : 170,1 Hz (σ = 26,4 Hz). On observe là une relation de cause à conséquence entre l'apparition du dysfonctionnement et la voix de celui qui le constate. Les valeurs de F_0 restent d'ailleurs relativement élevées pour la phrase qu'il prononce en suivant (voyelles 161 à 167).

Après cet épisode inquiétant, la fréquence fondamentale de l'OPL reprend des valeurs beaucoup plus basses des voyelles 168 à 229 malgré la voyelle isolée 216. Ces valeurs sont proches de celle des périodes précédentes équivalentes. La valeur moyenne vaut 125,0 Hz (σ = 21,4 Hz) pour les 62 voyelles en question. L'ensemble de ces trois périodes basses peut constituer l'image de la voix « au repos » ou « neutre » de l'OPL.

Il faut remarquer un contre exemple appartenant à cette période « neutre », celui de la phrase contenant les voyelles 189 à 197 pour lesquelles la F_0 est basse alors que l'OPL constate à haute voix et pour la deuxième fois le problème des commandes. Certes il ne s'agit plus d'un effet de surprise total mais on ne peut pas estimer que cet épisode appartient à un contexte « neutre ». C'est le cas pourtant du point de vue de la fréquence fondamentale. On observe là un effet nul du contexte sur les variations de F_0 .

À partir de la voyelle 230 et jusqu'à la voyelle 275, on est en présence d'une longue période de F_0 élevée dont la valeur moyenne vaut 179,6 Hz (σ = 50,2 Hz) pour les 46 voyelles qui y sont inscrites. La dispersion est importante. Le même type d'observation existe pour le CDB. Il s'agit d'une phase de discussion entre les membres d'équipage sur les causes possibles du problème.

De la voyelle 276 à la voyelle 312, la fréquence fondamentale reprend ses valeurs « neutres » même si quelques voyelles isolées font exception (300, 308 et 311). La moyenne est de 136,5 Hz (σ = 34,7 Hz) pour les 37 voyelles et de 129,2 Hz (σ = 24,8 Hz) sans les trois voyelles précédentes. Les pilotes sont toujours dans leur période de discussion. On retrouve une hausse sensible de F_0 pour la phrase des voyelles 313 à 323 (F_0 moyenne = 206,5 Hz, σ = 23,3 Hz) qui fait partie de cette discussion. Puis apparaît une zone de F_0 très basse suivie à nouveau de valeurs très élevées (voyelles 350 à 357 avec F_0 moyenne = 184,4 Hz et σ = 28,4 Hz).

La phase de discussion sur le dysfonctionnement se caractérise pour l'OPL par des hausses et des baisses successives de la fréquence fondamentale. Une fois la décision de décollage prise on retrouve une zone de F_0 « neutre » avant les derniers mots prononcés. Pour

les voyelles 358 à 396, la F_0 moyenne vaut en effet 134,12 Hz (σ = 29,07 Hz) pour les 39 voyelles segmentées.

Pour les derniers phonèmes prononcés alors que le problème se manifeste à nouveau en vol cette fois-ci, la hausse de la fréquence fondamentale s'accompagne d'une grande dispersion (de même pour le CDB) : F_0 moyenne = 171,2 Hz (σ = 52,3 Hz).

Ce type d'analyse a aussi été conduit pour le CDB et l'INGénieur navigant. Pour le premier, on retrouve les mêmes observations que pour l'OPL. En revanche, pour l'ING les fluctuations de F_0 sont très nettement moins présentes. L'hypothèse émise est que d'une part il est moins en prise directe avec le pilotage de l'avion que ses collègues et qu'il est donc probablement enclin à un moins grand degré de stress ou d'inquiétude.

Pour tous les autres paramètres du plan dynamique et du plan des phases, il n'a pas été possible de rapprocher leurs variations respectives aussi près du contexte. Certains éléments sont ponctuellement apparus comme une brusque diminution de la fréquence de modulation des maxima d'amplitude à la fin du vol pour le CDB (§ 3.5.4 page 101). Du point de vue spectral, les conclusions sont les mêmes. Mais la forme des enveloppes spectrales peut parfois conduire à des valeurs de paramètres localement inhabituelles.

Ces modifications de la distribution de l'énergie dans le spectre avaient été mises en évidence pour la thèse de Doctorat en comparant les distributions cumulées du niveau sonore. Par la suite, elles ont donné lieu à la définition de la fréquence d'équilibre spectral (§ 3.5.2 page 87) et aux mesures de centre de gravité spectral et de pente.

Lors des travaux de thèse, les distributions cumulées avaient été comparées pour une fréquence fixe de partage du spectre en une bande basse (BF) et une bande haute fréquence (HF). La première couvrait la zone des deux premiers formants et la seconde la partie complémentaire jusqu'à 5000 Hz. L'écart entre les deux distributions cumulées présentait un lien avec le contexte émotionnel. Les perturbations émotionnelles provoquaient aussi un accroissement de la fréquence du second formant.

Avec la fréquence d'équilibre spectral, qui peut varier d'un spectre à l'autre puisqu'elle est celle pour laquelle les deux distributions cumulées BF et HF qu'elle définit sont les plus proches, il n'a pas été observé de résultat significatif pour l'analyse des conversations de l'enregistreur phonique. Il en est de même pour le centre de gravité spectral.

Seules les variations de la pente spectrale des [a] de l'OPL peuvent présenter un lien avec le contexte.

Comme l'illustre la figure 3.1.1.2, elle décroit jusqu'à la voyelle 77 puis augmente jusqu'à la fin du vol. La voyelle 77 qui possède la pente la plus forte en valeur absolue, donc le spectre le moins plat, appartient à la phrase où le locuteur s'interroge sur la possibilité de voir par un hublot un quelconque élément de structure pouvant expliquer l'origine du problème de commande.

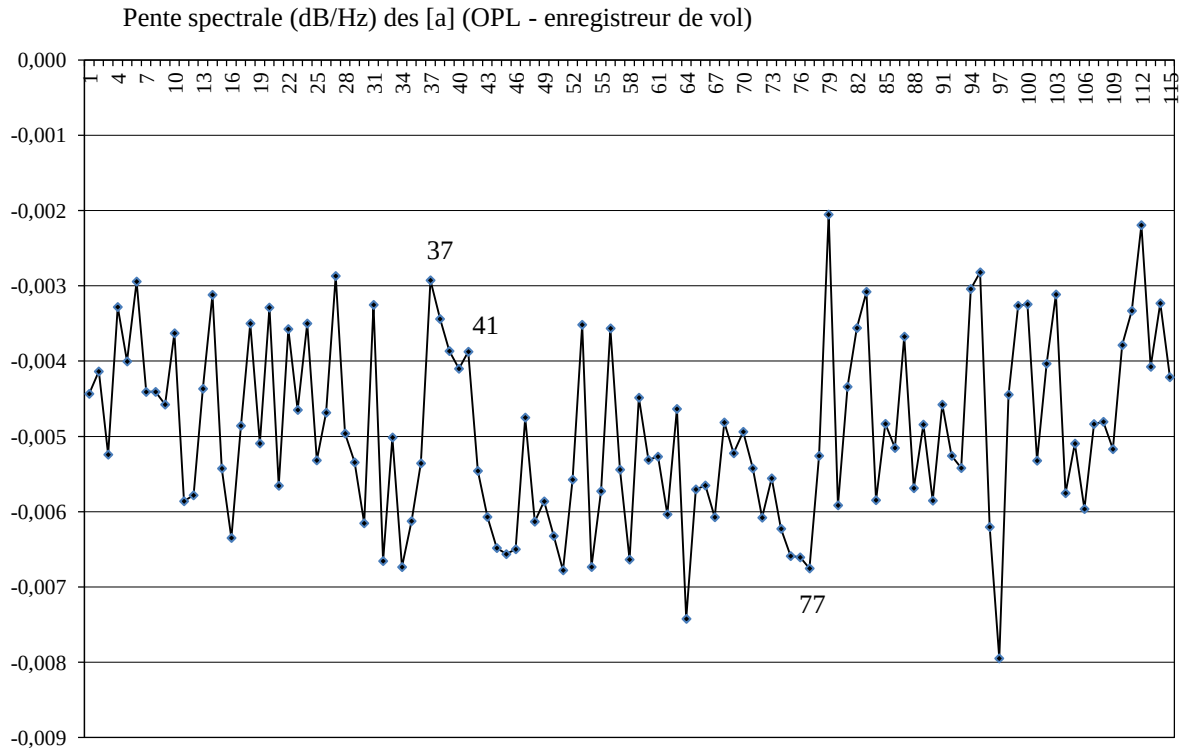


Figure 3.1.1.2 : Pente spectrale (dB/Hz) des [a] prononcés par l'OPL (enregistreur phonique).

À partir de cet instant, les pentes augmentent, traduisant un relèvement relatif du niveau des fréquences élevées. Le niveau de bruit de fond ne présente ni de hausse particulière ni de modification quelconque à partir de ce moment. Il est donc probable que les variations puissent être liées à l'état émotionnel du locuteur.

Mais l'interprétation est délicate car les variations de la pente spectrale peuvent être lues comme possédant deux sens de variation. Si on opte pour l'hypothèse que la dernière partie du vol est la plus stressante, alors l'accroissement de la pente en serait le reflet acoustique, et le changement de pente en deviendrait l'origine. On serait donc amené à considérer que les éléments de stress potentiels qui précèdent ne sont pas suffisants pour provoquer de tels changements. Peut-être y-a-t-il cumul d'un effet global indiqué par les deux sens de variation, avec des effets locaux comme celui des cinq voyelles consécutives 37 à 41 (figure 3.1.1.2) ?

Les faibles pentes (en valeur absolue) sont le signe de spectres plus plats, signes de perturbations du locuteur, conformément à la tendance générale observée pour ce paramètre dans la littérature, ce qui corroborerait les hypothèses formulées.

Les valeurs de début de vol, qui sont aussi importantes que celles de la fin, seraient alors liées à la tension qui précédait l'atterrissage dans des conditions de vol délicates. La décroissance traduirait par la suite un retour à la normale avant la hausse finale.

L'interprétation des variations est délicate. Elles sont développées pour témoigner de la difficulté de l'analyse des cas réels. Néanmoins ce cas est unique, car il fournit de nombreux éléments de conversation, pour trois locuteurs, avec plusieurs apparitions

consécutives de pannes et avec une excellente qualité d'enregistrement. Il a été choisi pour ces raisons et demeurera un élément de référence pour la poursuite des travaux, comme support de recherche lui-même ou pour tester la sensibilité des paramètres acoustiques.

Il est le siège de modifications vocales prosodiques et spectrales. En excluant les éléments lexicaux de la fin du vol, elles mettent en évidence la problématique de l'interprétation des résultats dans ces types de situation où les perturbations psychophysiologiques sont clairement présentes mais difficiles à localiser et à identifier.

3.1.2. Vol long-courrier avec incidents provoqués (partenaires : Services Techniques de la Navigation Aérienne - Airbus) 1993 [27]; L'étude a fait l'objet du co-encadrement d'un étudiant en thèse de Doctorat (B.Gramatica 1996).

Pour la certification en vol avec un équipage réduit à deux pilotes au lieu de trois personnels navigants, des vols long-courriers ont eu lieu avec création de pannes par les ingénieurs navigants. Les conversations de l'équipage ont été enregistrées pour étudier la corrélation entre le stress provoqué, les caractéristiques acoustiques de la voix et l'estimation de la charge de travail par des indicateurs médicaux mesurés par les partenaires de l'étude.

La société Airbus dispose d'un modèle statistique d'estimation de la charge de travail d'un pilote (« Airbus Workload Measure AWM » mis au point lors des certifications des Airbus A300, 310 et 320). Il prend en compte des paramètres de vol de l'avion, des paramètres médicaux, comme le rythme cardiaque, ainsi que des informations subjectives sur le contexte opérationnel fournies par les pilotes présents dans le poste de pilotage et complétées par des observations ergonomiques de spécialistes à bord.

Avec les enregistrements des conversations en vol, l'objectif était d'observer les variations de la fréquence fondamentale F_0 corollairement à celles du modèle AWM.

Il opère lors des situations émotionnelles qui suivent les pannes provoquées.

Lorsque les incidents surviennent, ils se traduisent par un accroissement des opérations à effecteur pour le pilotage et donc de la charge de travail. Ici, le champ émotionnel et le champ de l'activité se mêlent. Le terme charge de travail est un indicateur de l'activité du sujet.

Pour un des vols effectués on note une augmentation conjointe de F_0 et de la charge de travail.

La figure 3.1.2.1 montre des variations importantes de F_0 pour le pilote concerné (CM1) entre la période de repos choisie durant les échanges verbaux pendant la mise en route des moteurs (27 voyelles, F_0 moyenne = 138,0 Hz, σ = 10,4 Hz) et les périodes 1 et 3.

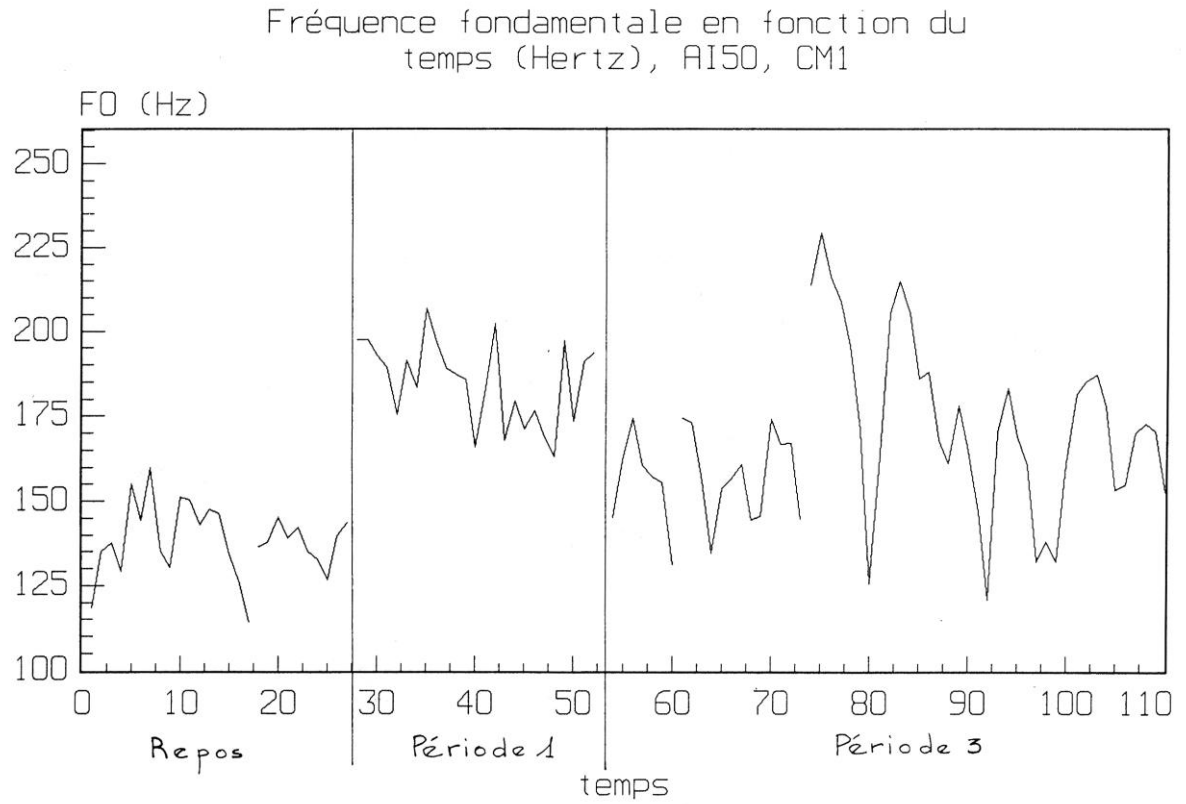


Figure 3.1.2.1 : Évolution de la fréquence fondamentale du pilote CM1 au cours du vol AI50 entre la période de repos et les périodes 1 et 3 [27].

La période 1 est une vérification de la procédure d'allumage des moteurs au cours de laquelle le pilote CM1 s'adresse à son copilote avec une voix interrogative. L'accroissement important de F_0 (25 voyelles, F_0 moyenne = 185,0 Hz, σ = 12,1 Hz) est due à la nature interrogative de la prononciation.

Ce n'est pas le cas de la période 3 qui suit la mise en panne d'un moteur en vol (68 voyelles, F_0 moyenne = 165,7 Hz, σ = 22,3 Hz). La hausse par rapport à la période de repos est statistiquement significative et on note une dispersion plus élevée.

Parallèlement, la charge de travail passe de 2 à 4 sous la forme d'un pic à 8h05 GMT, heure à laquelle a eu lieu la panne moteur (période 3 sur la figure 3.1.2.1 et « Eng. 2 Fail » sur la figure 3.1.2.2).

L'apparition de la modification vocale est légèrement différée par rapport à l'incident.

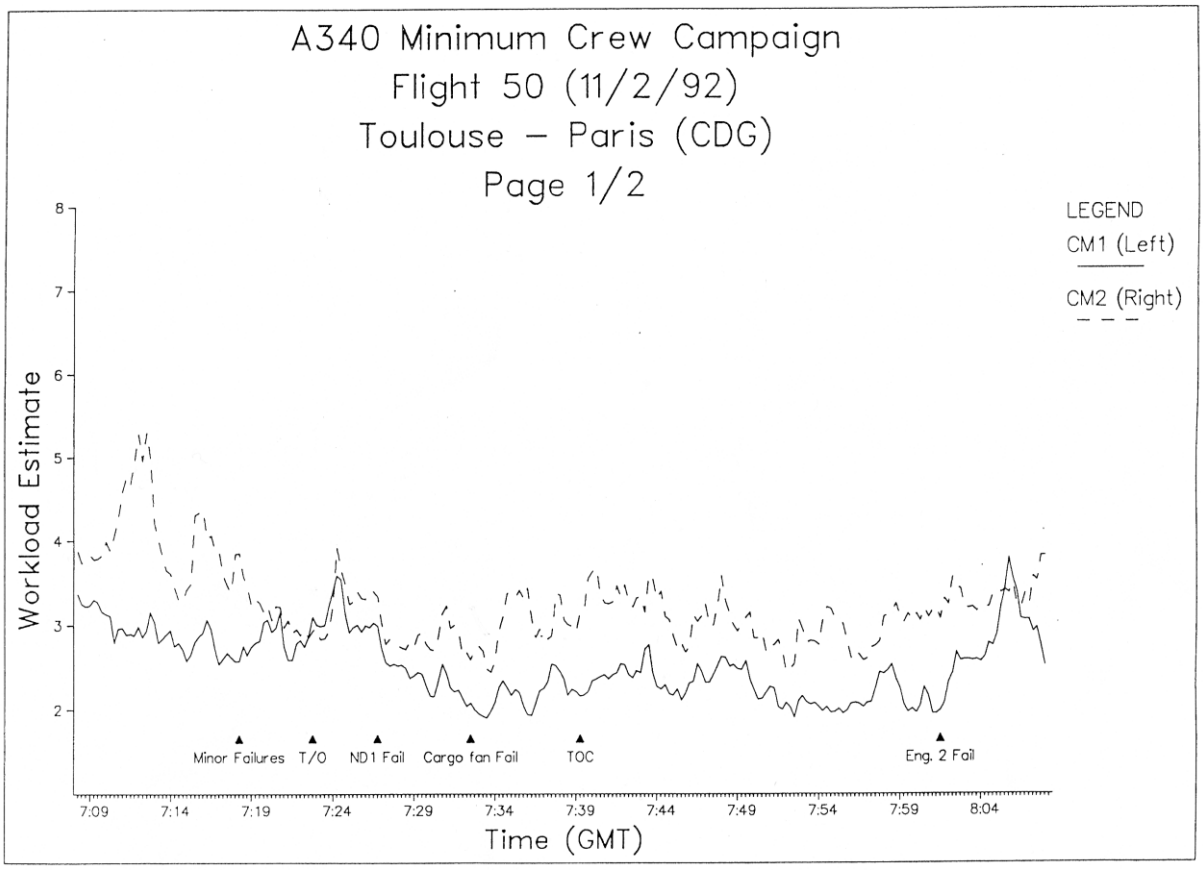


Figure 3.1.2.2 : Évolution de la charge de travail au cours du temps pour les deux pilotes [27].

Lors des autres vols de l'expérimentation, il a pu être observé de nombreux cas où la F_0 moyenne augmentait sans modification de la charge de travail.

Ces modifications sont dues à la teneur du discours et au contexte dans lequel sont plongés les locuteurs. Par exemple, lorsqu'un pilote essaie d'en convaincre un autre, ou lors de communications radio ou encore lorsqu'un d'entre eux manifeste des difficultés à s'exprimer dans une autre langue que sa langue maternelle.

Il a également été observé des situations où la fréquence fondamentale moyenne reste stable alors que la charge de travail augmente. Ceci a eu lieu lors de situations aéronautiques graves et peut s'expliquer par la dissipation du stress lorsque les pilotes ont pris la parole après avoir réalisé d'importantes et rapides manipulations des instruments de vol et rétabli la situation.

Il apparait que la modification de l'écart-type σ des distributions de la fréquence fondamentale est en relation avec l'apparition des pannes. Lorsque la dispersion est importante, typiquement avec un écart-type supérieur à 20 Hz, il s'agit dans la plupart des cas de phases où un incident est survenu.

Enfin, le résultat le plus singulier concerne un des vols de l'expérimentation où la charge de travail et le jitter suivent la même évolution (figure 3.1.2.2). Il s'agit du jitter des fréquences fondamentales moyennes par phonème.

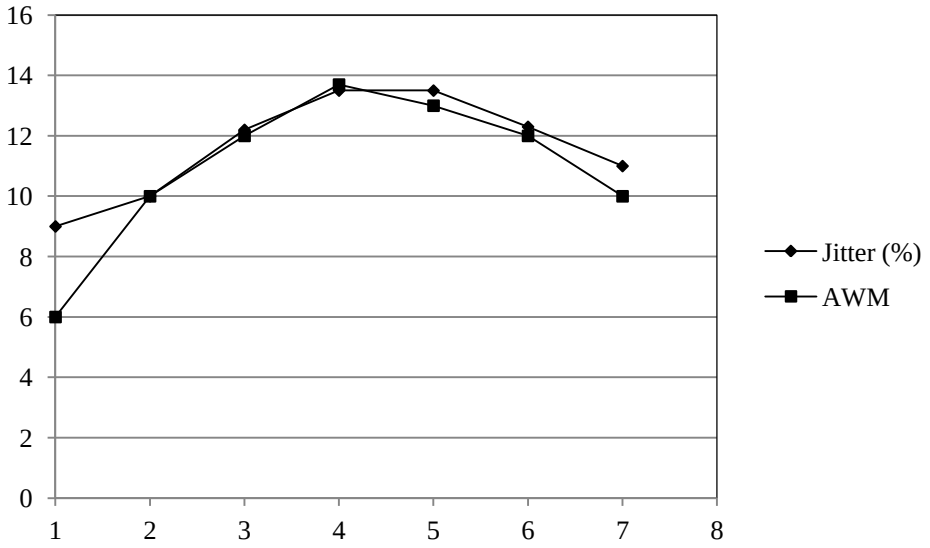


Figure 3.1.2.2 : Jitter (en %) et charge de travail (AWM sans unité) en fonction des périodes de vol (la première est celle de repos) (thèse B.Gramatica, 1996).

Une corrélation stricte, au sens mathématique, entre la charge de travail du modèle et la fréquence fondamentale n'est pas constatée. La fréquence fondamentale porte d'autres informations, notamment liées au style de la prononciation, que celles des effets probables de la charge de travail. L'estimation médico-subjective de la charge de travail igore l'absence de parole. Les conclusions de l'étude stipulent que d'autres paramètres acoustiques sont à mesurer pour, autant que possible, extraire les seuls effets de la cause perturbatrice. L'évidence qu'une analyse vocale n'est possible que s'il y a prise de parole prend son sens dans ces cas réels de pannes où le pilote peut s'exprimer tardivement après l'apparition de l'incident.

Du point de vue spectral, la fréquence d'équilibre spectral et l'aire à cette fréquence ont été modifiées par l'état de stress. Dans ces situations, la première prend des valeurs en moyenne fréquence (cf §3.5.2 page 87) mais des voyelles prononcées au repos le vérifient également. Il n'est donc pas possible de conclure sur les effets spectraux du phénomène.

L'expérience a servi de support pour l'analyse des consonnes. L'importante variabilité de leurs caractéristiques acoustiques n'a pas permis de les mettre en relation avec les perturbations du locuteur (thèse B.Gramatica, 1996).

3.1.3. Simulateur de vol avec pannes provoquées (partenaires Aéroformation et Base Opérationnelle Mobile AéroPortée) 1990 à 1992 [7], [9]; Pour la première des deux expérimentations, le travail a fait l'objet du co-encadrement d'un D.E.A en acoustique (B.Gramatica 1992).

Le simulateur a permis de construire un protocole expérimental pour provoquer en situation contrôlée divers incidents de vol et en analyser les conséquences vocales.

Un élève-pilote a effectué 1h30 de vol simulé d'hélicoptère au cours duquel il a du régler ou s'accommoder de plus d'une dizaine de pannes. Il a également fait deux erreurs de pilotage.

Les résultats des analyses vocales montrent des hausses significatives de la fréquence fondamentale pour plus de la moitié des pannes et pour les deux erreurs commises.

La référence dite de « repos » correspond à une moyenne de F_0 pour 83 voyelles prononcées avant le vol.

Mais la significativité des augmentations n'est pas confirmée lorsque les prononciations après les incidents sont comparées à celles d'avant l'apparition des pannes.

On observe toutefois une hausse graduelle de F_0 au cours de la séance qui peut résulter de l'accumulation importante de dysfonctionnements de plus en plus graves jusqu'à la fin du vol. Dans ces conditions, on peut émettre l'hypothèse que le pilote se trouve plongé dans un état presque permanent de stress et de tension nerveuse qui empêche tout retour à la normale des valeurs de fréquence fondamentale. Ceci est d'autant plus vrai que la durée du vol est relativement courte eut égard au nombre de pannes provoquées.

La dispersion des valeurs de F_0 n'est pas sensiblement modifiée au cours de l'expérimentation.

Pour une seconde expérimentation avec deux pilotes instructeurs à bord d'un simulateur de vol d'Airbus, des hausses de fréquence fondamentale ont été mesurées après l'apparition d'incidents.

La figure 3.1.3.1 illustre le phénomène : toutes les voyelles segmentées après tous les incidents sont prises en compte pour la construction de l'histogramme lissé des valeurs de F_0 noté « stress » (de même pour la situation de « repos »).

La fréquence fondamentale croit dans un rapport 1,4 entre la situation de repos et celle de stress. La référence de repos a été choisie en début de vol avant l'apparition des premières pannes.

Ce type de schéma ne signifie pas que toutes les pannes ont systématiquement occasionné une hausse de F_0. Celles qui sont dans ce cas le sont à la fois par rapport à la phase de repos et par rapport aux prononciations qui précèdent l'apparition de l'incident. À titre d'exemple, la figure 3.1.3.2 détaille les valeurs obtenues et leur intervalle de confiance à 95% pour un des deux pilotes.

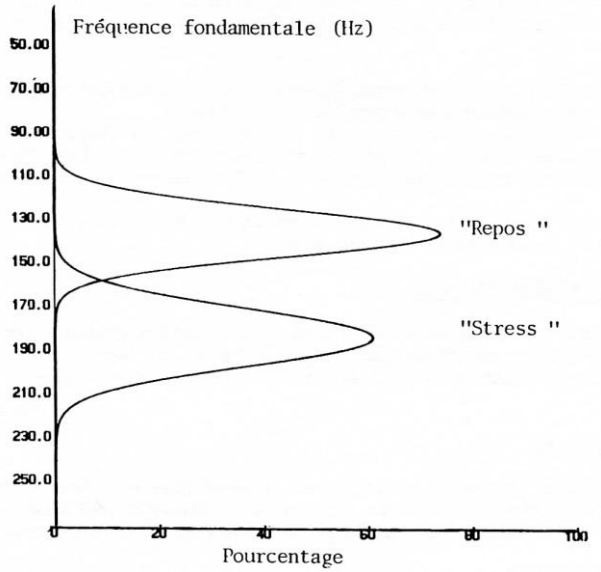


Figure 3.1.3.1 : Fréquence fondamentale moyenne pour les deux locuteurs et pour toutes les phases de stress [7].

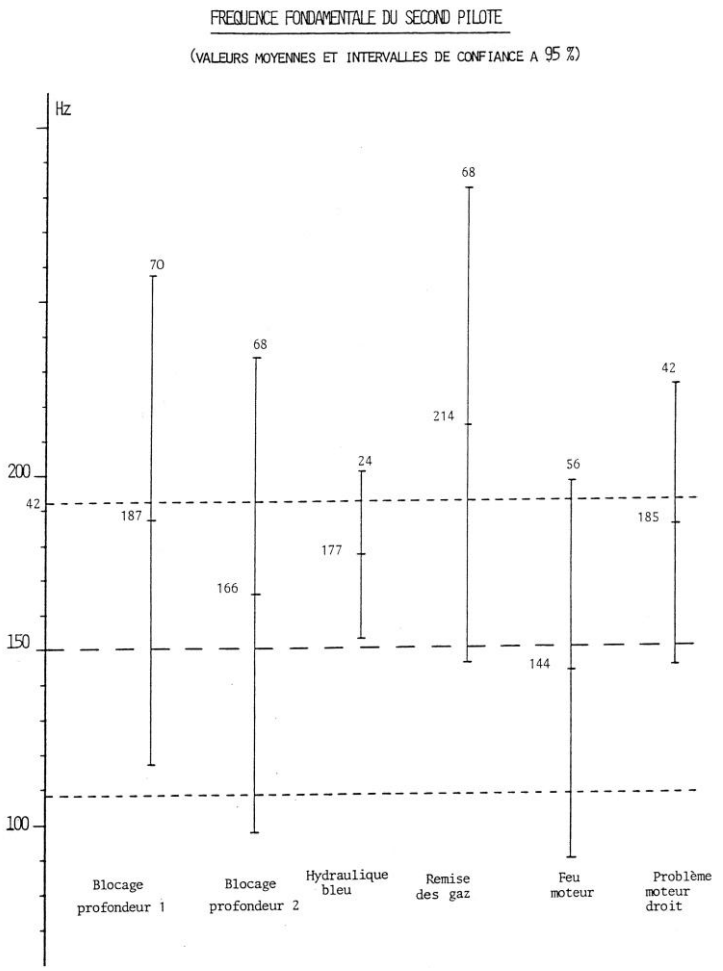


Figure 3.1.3.2 : Fréquence fondamentale moyenne et intervalle de confiance à 95% pour différents incidents de vol simulés. [26]

Corollairement, la dispersion des valeurs subit d'importantes variations.

La fréquence, l'amplitude et la largeur de bande des trois premiers formants des voyelles les plus représentées dans les segments extraits pour la mesure de F_0 ont été mesurées. Aucune tendance particulière n'a pu être mise en évidence lorsque les incidents apparaissent.

Le simulateur confirme sa prédisposition à induire, comme dans les cas réels, des modifications vocales. Les incidents, source de stress et de tension nerveuse, provoquent des hausses de la fréquence fondamentale.

Bien que regroupés sous l'intitulé « voix et émotion », les travaux décrits dans ce paragraphe 3.1 ne concernent que leur lien dans le domaine aéronautique. Ils sont à rapprocher des études similaires qui ont été recensées pour le secteur des transports (§2.1 page 8). S'agissant d'expressions émotionnelles, ils sont aussi à comparer avec la revue des connaissances sur la voix et l'émotion (§2.2 page 15).

Les observations qui découlent de la mise en parallèle des études sont les suivantes.

- Les investigations acoustiques menées sur les conversations d'un enregisteur de vol sont rares. Bien que les évènements ne soient pas comparables, la hausse et la modification du contour temporel de la fréquence fondamentale sont identiques. La contribution a été de tenter d'interpréter les variations, au plus près du contexte de la situation. L'apport a été également d'engager la mesure d'autres paramètres, et d'aboutir à observer des modifications locales et individuelles (fréquence d'équilibre spectral, pente du spectre, modulation d'amplitude des maxima du signal temporel).

- En conditions non dramatiques de pannes réelles provoquées ou d'incidents simulés, les travaux entrepris sont moins diversifiés que leurs homologues parce qu'ils n'ont traités que de la fréquence fondamentale des voyelles. Ils sont intervenus à une période où les mesures d'autres caractéristiques n'étaient pas mises au point et où il s'est agit d'explorer les propriétés des consonnes (thèse B.Gramatica 1996). Ces travaux ont d'ailleurs aboutit à recentrer les recherches sur les voyelles, faute de résultats suffisamment stables pour les consonnes.

- Hors du secteur des transports, les comparaisons avec les études générales sur l'émotion n'ont pas lieu d'être au cas par cas, car les processus émotionnels et les expériences sont nettemment différents. Il est par contre possible de conclure que des signes acoustiques communs existent autour des variations de la fréquence fondamentale.

- L'universalité des variations de la fréquence fondamentale sous l'effet de toutes les conditions expérimentales lui donne un caractère témoin mais rend complexe l'éventualité d'une classification typologique sans l'apport d'autres paramètres.

3.2. Fatigue

Des études en vol réel et en conditions contrôlées ont été effectuées. Sous l'appellation fatigue sont présentées plusieurs expérimentations :

- sur une série de vols court-courriers consécutifs,
- sur la conduite automobile,
- sur l'inertie du sommeil,
- sur la fatigue physique.

Le terme de fatigue est sans doute impropre pour désigner les conséquences des phénomènes perturbateurs de ces diverses situations. Néanmoins, ils sont tous distincts de ceux qui induisent de l'émotion ou du stress. Ils sont inscrits sous le terme générique de fatigue. On ne pourra toutefois pas en généraliser les résultats pour la détection ou la caractérisation de la fatigue dans son acception la plus générale. Les conclusions de chaque expérimentation lui sont propres.

3.2.1. Vols Court-courriers consécutifs (partenaires : Bureau d'Enquêtes et d'Analyses, Air France), 2008 à 2010, [5], [30], [32], [33].

Il s'est agit d'évaluer la fatigue d'un équipage au cours d'une journée de vols court-courriers avec plusieurs rotations sur le territoire français.

Les conditions de travail des pilotes peuvent induire de la fatigue due au nombre de vols effectués quotidiennement avec le même appareil accompagnés de réveils très matinaux.

Un équipage de deux pilotes (commandant de bord CDB et officier pilote de ligne OPL) a été enregistré à chaque escale d'une journée de vol. Il lui a été demandé de prononcer cinq phrases choisies pour leur similitude avec le vocabulaire aéronautique, et comportant toutes l'indicatif de l'avion.

L'objectif est d'examiner si des paramètres acoustiques peuvent présenter des évolutions significatives au cours de la journée.

La palette des caractéristiques acoustiques mesurées est large puisqu'elle couvre les plan dynamique, des phases et spectral.

Trois enregistrements ont été analysés : au départ à 6h, à 8h au départ de la première escale et à 15h à l'arrivée du dernier vol.

Le choix de 8h comme enregistrement intermédiaire résulte du fait que les pilotes ont exprimé des signes de lassitude à ce moment là (baillements).

Les résultats les plus remarquables concernent le paramètre DAL_{To}/T_0 (cf § 3.5.3 pour sa définition page 95), les distances spectrales et le plus grand exposant de Lyapunov.

Sur la figure 3.2.1.1 on remarque une hausse des valeurs de DAL_{To}/T_0 à 8h (Recording 2) par rapport à celles de 6h (Recording 1) et à celles de 15h (Recording 3).

La hausse peut aussi être observée sur le tracé de l'évolution chronologique de la fréquence fondamentale comme le montre la figure 3.2.1.2, mais avec une augmentation globale moins nette.

Le paramètre DAL_{To}/T_0 possède la propriété de mettre en évidence, de manière plus sensible, de faibles variations de la fréquence fondamentale en relation avec l'apparition de signes de fatigue.

D'ailleurs, si les valeurs de la fréquence fondamentale en fin de journée sont peu différentes de celles du matin, celles du paramètre DAL_{To}/T_0 apparaissent en revanche plus élevées (cas du premier pilote seulement).

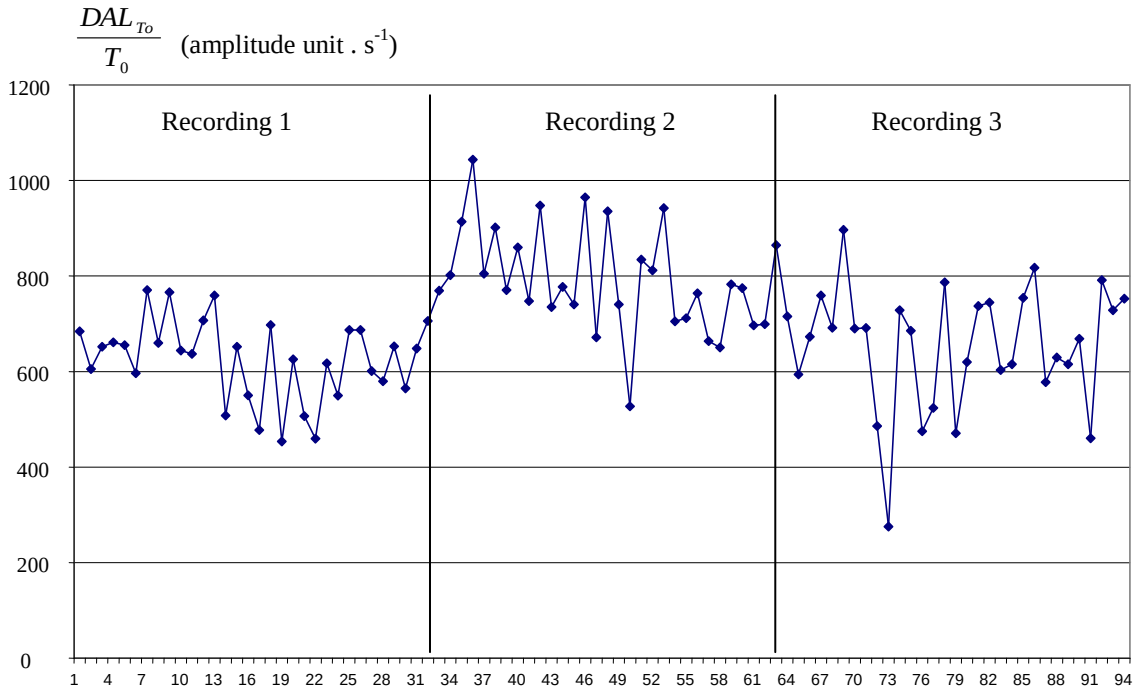


Figure 3.2.1.1 : Variations du paramètre DAL_{To}/T_0 pour le premier pilote en fonction du n° d'ordre des voyelles segmentées [5].

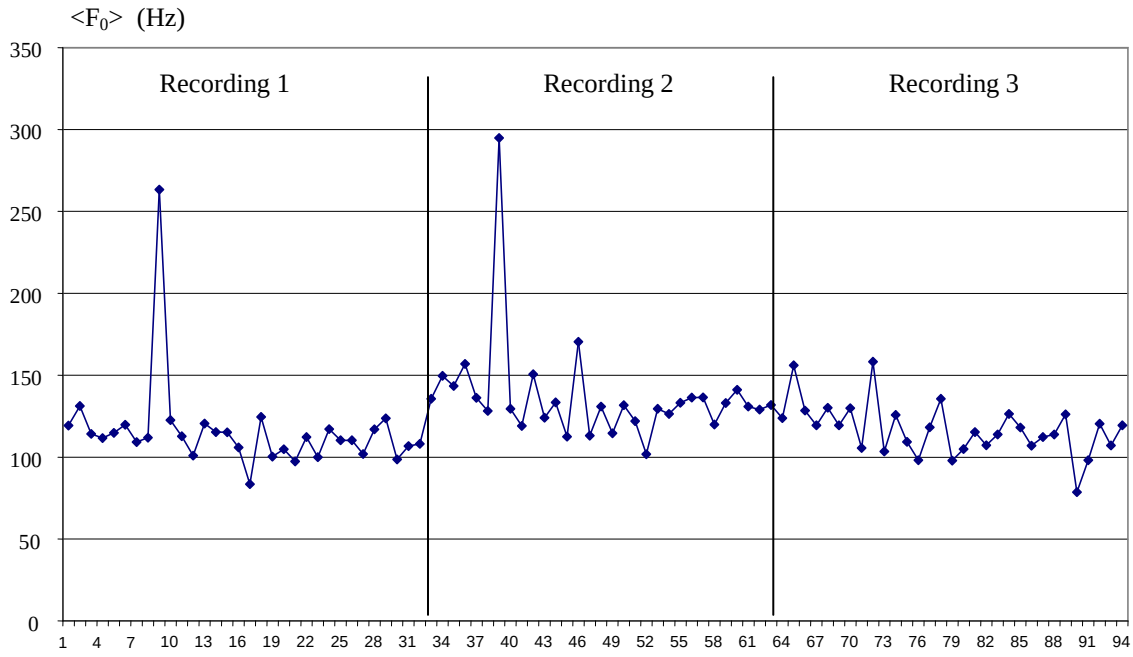


Figure 3.2.1.2 : Variations de la fréquence fondamentale moyenne pour le premier pilote en fonction du n° d'ordre des voyelles segmentées [5, 32].

[a]	Pilot 1	Pilot 2
Rec. 1 – Rec. 2	5,51	5,47
Rec. 2 – Rec. 3	3,85	2,15
Rec. 1 – Rec. 3	3,52	5,92

[e]	Pilot 1	Pilot 2
Rec. 1 – Rec. 2	5,10	7,76
Rec. 2 – Rec. 3	4,61	2,66
Rec. 1 – Rec. 3	2,41	6,12

[ε]	Pilot 1	Pilot 2
Rec. 1 – Rec. 2	5,64	5,59
Rec. 2 – Rec. 3	5,70	4,02
Rec. 1 – Rec. 3	3,96	3,69

[i]	Pilot 1	Pilot 2
Rec. 1 – Rec. 2	6,88	5,99
Rec. 2 – Rec. 3	6,14	3,67
Rec. 1 – Rec. 3	2,56	4,02

[o]	Pilot 1	Pilot 2
Rec. 1 – Rec. 2	3,70	2,34
Rec. 2 – Rec. 3	3,43	1,63
Rec. 1 – Rec. 3	2,44	3,17

Tableau 3.2.1.1: Distances spectrales entre les périodes d'enregistrement (Rec.) pour chaque type de voyelle et pour les deux pilotes [5, 32].

Concernant les distances spectrales, on observe que quelle que soit la voyelle considérée [a], [e], [ε], [i], [o] elles sont toujours plus faibles entre le matin et la fin de journée qu'entre le matin et l'escale de 8h et ce pour les deux pilotes (sauf pour les [a] et [e] du second pilote) (Tableau 3.2.1.1).

Le tracé d'une régression linéaire sur les variations chronologiques du plus grand exposant de Lyapunov fait apparaître une hausse générale (figure 3.2.1.3).

Il s'agit des voyelles prononcées par le second pilote. Pour le premier la pente est quasi nulle. En revanche, pour les deux pilotes c'est à 8h que la moyenne des plus grands exposants de Lyapunov calculés pour chaque voyelle est la plus élevée. Pour 6h et 15h sa valeur moyenne est plus faible (Tableau 3.2.1.2).

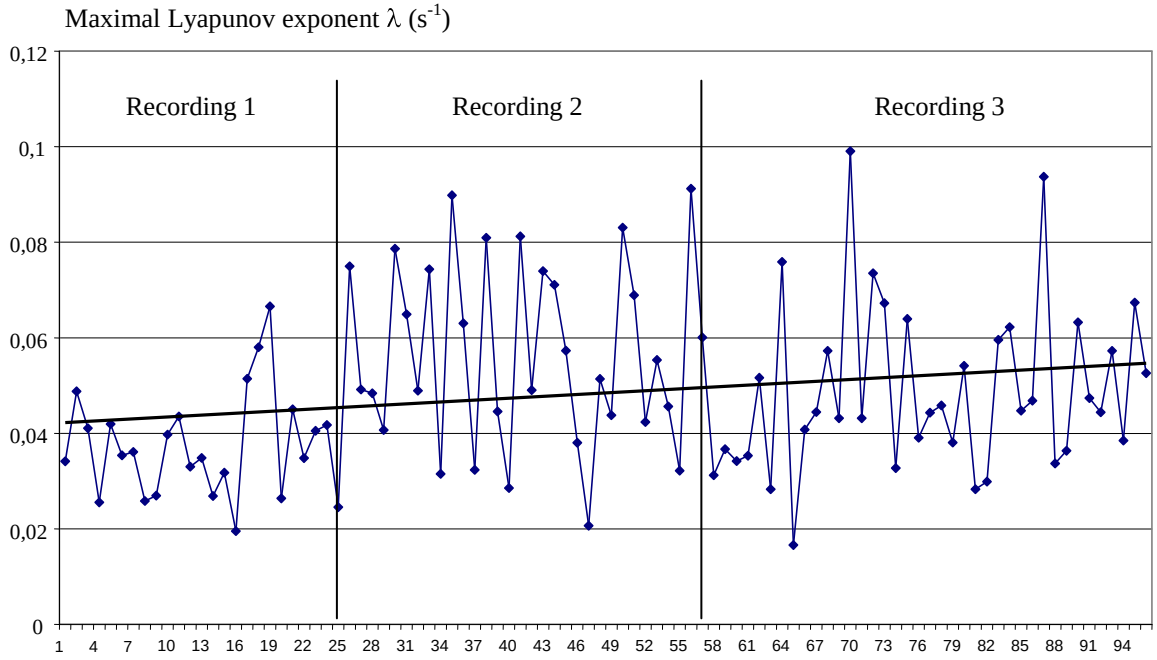


Figure 3.2.1.3 : Variation du plus grand exposant de Lyapunov en fonction du n° d'ordre des voyelles segmentées pour le second pilote [5, 33].

	λ moyen pour le CDB (σ)	λ moyen pour l'OPL (σ)
« Orly 6h » (Rec.1)	0,0506 (0,0188)	0,0379 (0,0110)
« Pau » (Rec.2)	0,06005 (0,01653)	0,0558 (0,0198)
« Orly 15h » (Rec.3)	0,0504 (0,0180)	0,0487 (0,0176)

Tableau 3.2.1.2 : Exposant de Lyapunov moyen par périodes de vol pour chaque pilote [33].

En définitive, il apparaît que des caractéristiques acoustiques permettent de conclure qu'entre le début et la fin de la journée de travail des modifications vocales ont eu lieu.

La fatigue serait traduite vocalement par une hausse du plus grand exposant de Lyapunov et des modifications des distances spectrales.

Il semblerait aussi que lors de l'escale intermédiaire de 8h, les signes extérieurs de fatigue qui avaient pu être repérés par l'expérimentateur dans la cabine de pilotage, puissent trouver une traduction acoustique exprimée par le paramètre DAL_{To}/To et le plus grand exposant de Lyapunov. Ce sont peut-être là des effets dus à la mise en activité consécutivement à une privation de sommeil.

3.2.2. Conducteur Automobile (partenaire de l'étude : Bureau d'Enquêtes et d'Analyses), 2009, [23], [34].

L'objectif était de mettre en place une expérimentation de conduite automobile pour laquelle la charge de travail pouvait être comparable avec celle des pilotes de vols court-courriers quotidiens. Plus facile à mettre en œuvre que les expériences en vol, elle a permis de tester les paramètres acoustiques de l'analyse de la voix dans cette configuration.

Le conducteur locuteur a parcouru 700 km de conduite diurne. Toutes les deux heures environ il prononçait les cinq mêmes phrases que celles utilisées pour l'enregistrement de l'expérience des vols court-courriers (§ 3.2.1 page 63). Les voyelles segmentées pour l'analyse appartenaient à ces phrases.

Les résultats ont établi que certains paramètres peuvent être en relation avec la fatigue de la conduite automobile.

En fin de journée le shimmer a chuté sensiblement (figure 3.2.2.1).

Le centre de gravité de plusieurs voyelles a augmenté tout au long de la journée (figure 3.2.2.2). L'accroissement est de 26% pour les [i] (figure 3.2.2.2) et il est en moyenne de 6.2% pour les autres types de voyelles.

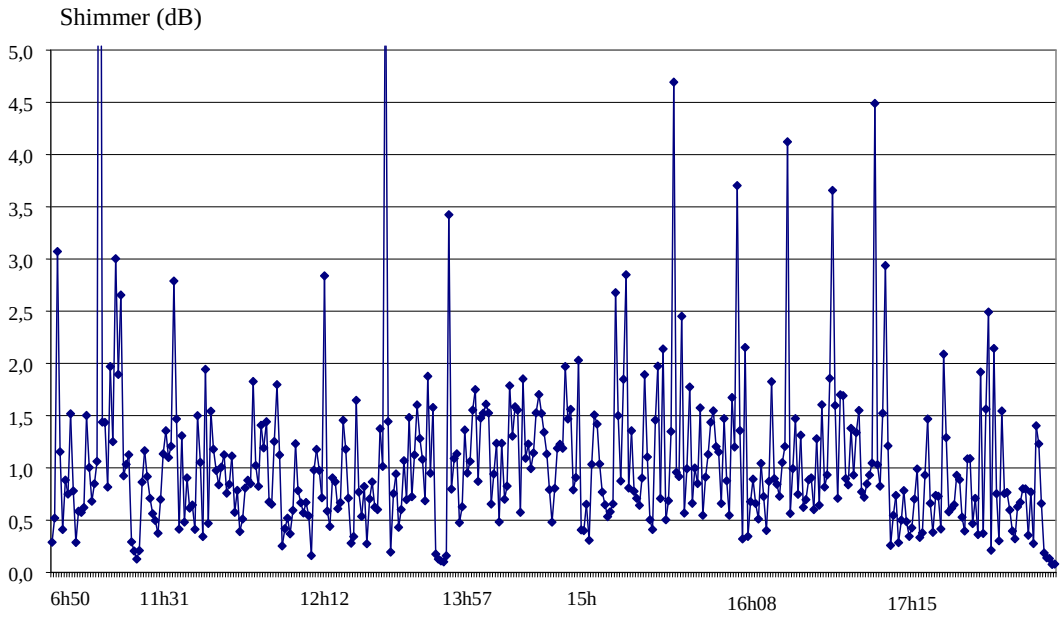


Figure 3.2.2.1. Évolution chronologique du shimmer (dB) au cours de la journée de conduite automobile [23, 34].

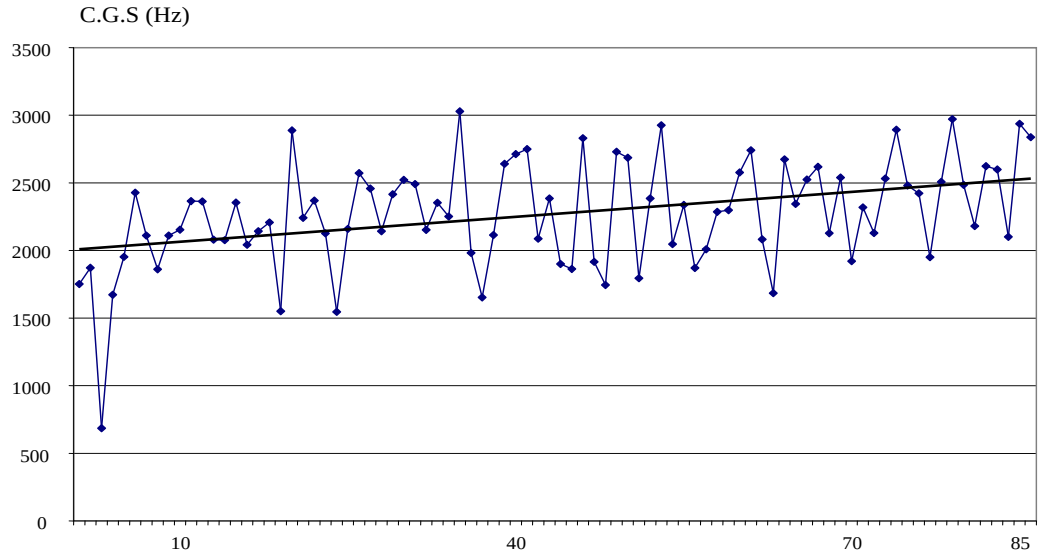


Figure 3.2.2.2. Évolution chronologique du centre de gravité spectral (C.G.S en Hz) pour les voyelles [i] au cours de la journée de conduite automobile [23, 34].

L'étude des distances spectrales a mis en évidence l'importance du choix de la référence du spectre moyen « au repos ». Les variations possèdent en effet un sens différent selon le choix effectué : en début de journée (et la veille) ou après deux heures de conduite.

La mesure du paramètre DAL_{To}/T_0 a permis, comme pour la fréquence fondamentale, mais de manière plus sensible, d'isoler les prononciations matinales de celles de la veille ou du reste de la journée. C'est un résultat analogue à celui obtenu pour l'expérience des vols court-courriers (§ 3.2.1 page 63).

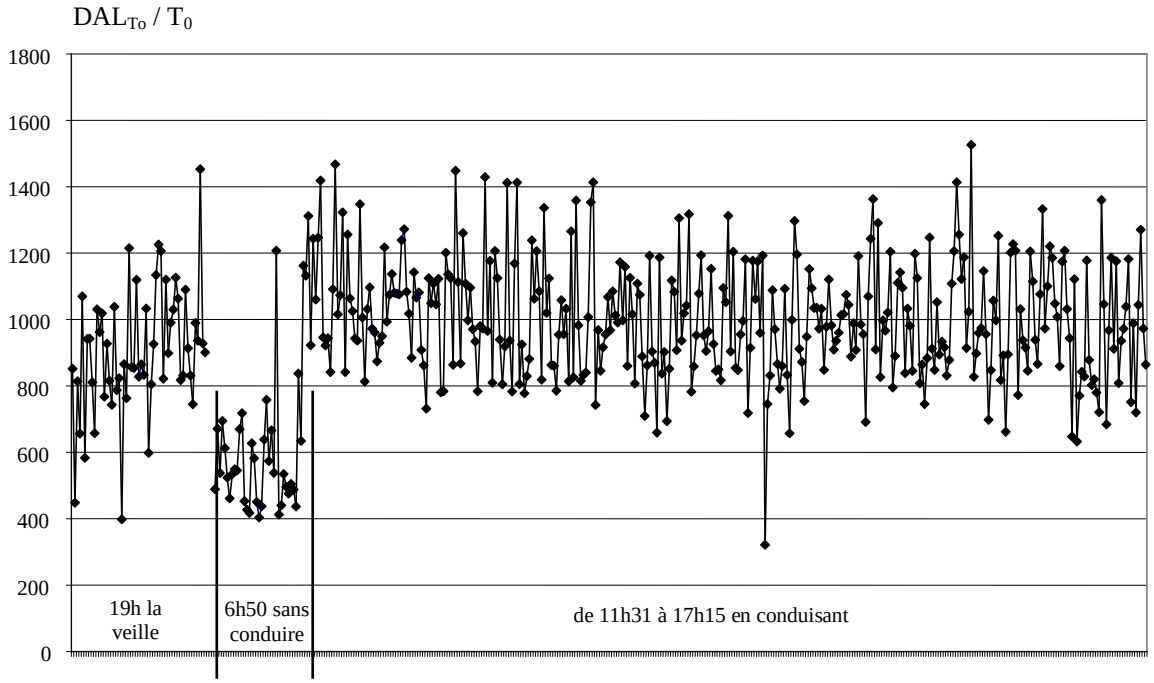


Figure 3.2.2.3 : Évolution du paramètre DAL_{To}/T_0 depuis la veille au soir de l'expérimentation (19h) jusqu'à la fin de la journée de conduite (17h15) [23, 34].

Le phénomène est certainement du à un effet de somnolence ou à une expression vocale avec une voix encore peu utilisée, très tôt le matin, avant le début de la conduite automobile.

Des mesures de débit phonatoire ont également montré qu'il décroît au cours de la journée d'environ 50 phonèmes par minute pour chacune des cinq phrases prononcées (figure 3.2.2.4). Leur durée moyenne augmente de quelques secondes entre le début et la fin de la journée. Sur la figure 3.2.2.4 le premier enregistrement est celui qui a été réalisé la veille à 19h, le second à 6h50 et le huitième à la fin de la journée (17h15).

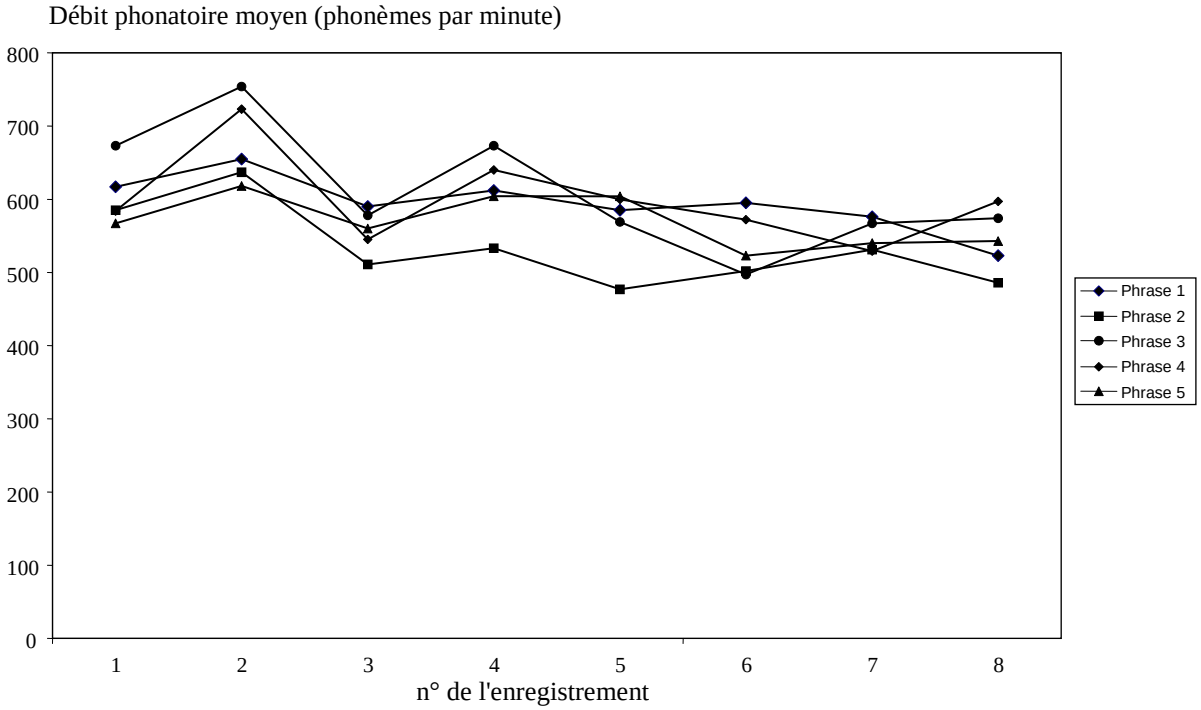


Figure 3.2.2.4 : Variation chronologique du débit phonatoire moyen (en phonèmes par minute) de chaque phrase [34].

Aucun autre paramètre acoustique ne présente de modifications notables.

Outre les résultats précédents, l'expérience souligne à nouveau l'importance du choix de l'échantillon de référence dit de « repos » pour l'interprétation des variations. Sur les variations de DAL_{To}/T_o (ou sur celles de la fréquence fondamentale) le choix de la période de repos quelques heures avant le début de la conduite (6h50 figure 3.2.3.3) n'est pas approprié : il pourrait provoquer une significativité des comparaisons erronée puisque cette période est elle-même l'expression d'une perturbation.

L'expérience de conduite automobile vient souligner la propriété du paramètre DAL_{To}/T_o d'exagérer les modifications relatives de la fréquence fondamentale. Avec l'appui des résultats de l'expérience des vols court-courriers, il semble offrir une plus grande dynamique de variation et des aptitudes à déceler les conséquences des fatigues matinales. Mais son sens de variation est différent dans les deux expériences. Il est donc prématuré de

statuer sur ses propriétés parce que le nombre de locuteurs est faible et parce que les causes perturbatrices ne sont pas directement comparables.

La longue conduite automobile a provoqué des effets acoustiques sur la voix du conducteur : ralentissement du débit phonatoire, chute du shimmer, hausse du centre de gravité spectral.

3.2.3. Inertie du Sommeil : enregistrements couplés avec le protocole expérimental de laboratoire mis en place pour étudier l'inertie du sommeil (partenaires Laboratoire d'Anthropologie Appliquée L.A.A, hôpital Henri Mondor et B.E.A), 2007 à 2010 [5], [21], [28], [31], [33];

Pour l'étude de la vigilance des équipages, le L.A.A a développé un protocole expérimental de laboratoire pour contrôler le sommeil qui a permis d'enregistrer la voix d'un pilote à son réveil lors de l'exécution de tâches aéronautiques. L'étude poursuivait le but d'identifier les caractéristiques non médicales, donc vocales, modifiées par l'inertie du sommeil.

L'expérience répond au souci d'évaluer les capacités d'un pilote à prendre les décisions puis à réaliser les actions appropriées suite à un réveil brutal du au déclenchement d'une alarme. Cela pourrait être le cas à l'apparition d'un incident en vol long-courrier.

Un locuteur, pilote expérimenté, a prononcé les cinq mêmes phrases issues du vocabulaire aéronautique avant (recording 1) puis après déjeuner (avant son endormissement) (recording 2). Il a été réveillé subitement avec un éclairage très puissant lorsque son encéphalogramme révélait qu'il était dans un sommeil profond. Il a du réaliser immédiatement après son réveil, sur un micro ordinateur, des activités psychomotrices en lien avec le pilotage. A l'issue de cette activité de quelques minutes il a prononcé à nouveau les cinq phrases (recording 3). Les voyelles analysées sont segmentées dans ces cinq phrases pour chacune des trois phases de l'expérimentation.

L'analyse des enregistrements montre que le jitter (figure 3.2.3.1) et la fréquence de modulation des maxima d'amplitude (cf §3.5.4 page 101) varient très sensiblement lors des prononciations après le réveil. Ces paramètres de la micro prosodie sont affectés par le phénomène tout comme le shimmer qui présente une légère hausse au réveil du locuteur (figure 3.2.3.2).

Le jitter vaut en moyenne 2.67 Hz (σ = 1.77 Hz) avant l'endormissement (recordings 1 et 2) puis 7.82 Hz (σ = 2.89 Hz) après le réveil (recording 3) [5,31]. Des allures similaires sont observées pour l'écart-type de la valeur moyenne de la fréquence fondamentale de chaque voyelle et de son coefficient de variation [31].

Pour le shimmer la hausse est moins nette. Il présente une valeur moyenne de 5.5 dB (σ = 6.9 dB) pour les deux premières phases puis une valeur moyenne de 9.75 dB, (σ = 9.4 dB) pour la dernière [5, 31].

Les résultats concernant la fréquence de modulation des maxima d'amplitude sont détaillés et discutés au paragraphe 3.5.4 (page 101) où ce paramètre est défini.

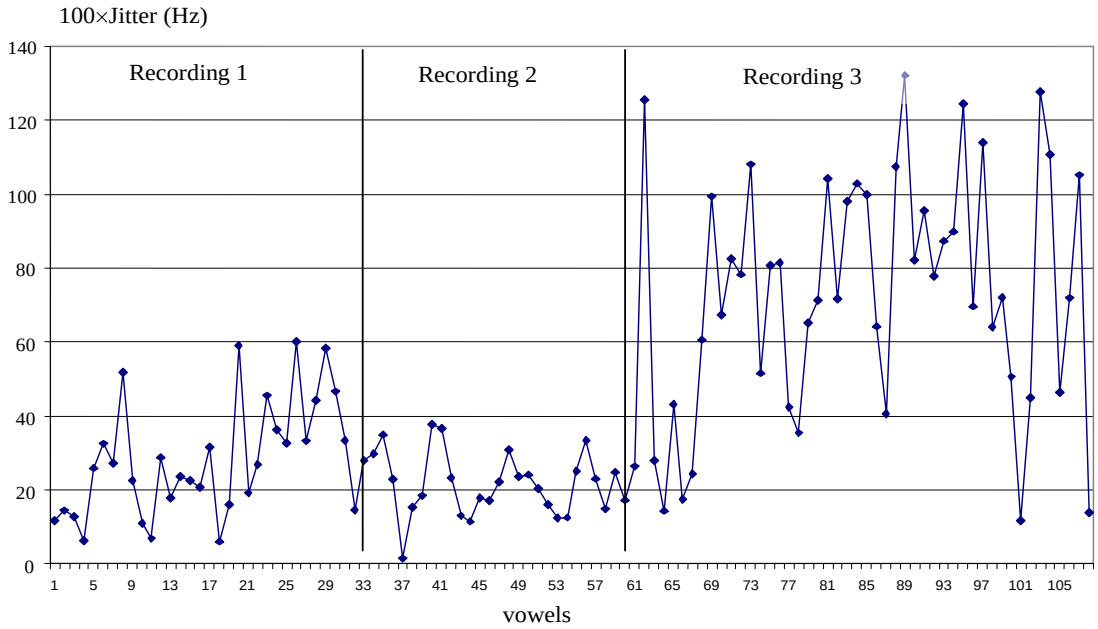


Figure 3.2.3.1 : Évolution du jitter au cours des trois phases de l'expérimentation sur l'inertie du sommeil en fonction du numéro d'ordre des voyelles : avant l'endormissement (recordings 1 et 2) et après le réveil (recording 3) [5, 31].

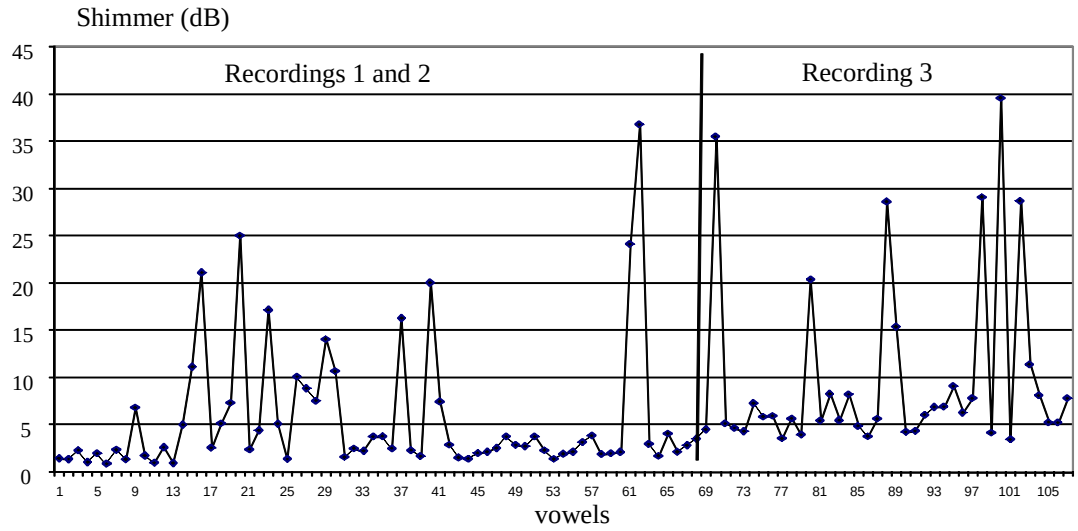


Figure 3.2.3.2 : Évolution du shimmer au cours des trois phases de l'expérimentation sur l'inertie du sommeil en fonction du numéro d'ordre des voyelles : avant l'endormissement (recordings 1 et 2) et après le réveil (recording 3) [5, 31].

La mesure du plus grand exposant de Lyapunov a également montré une tendance à la décroissance comme l'illustre la figure 3.2.3.3. Les valeurs en phase 3 sont significativement différentes de la moyenne de celles en phase 1 et 2.

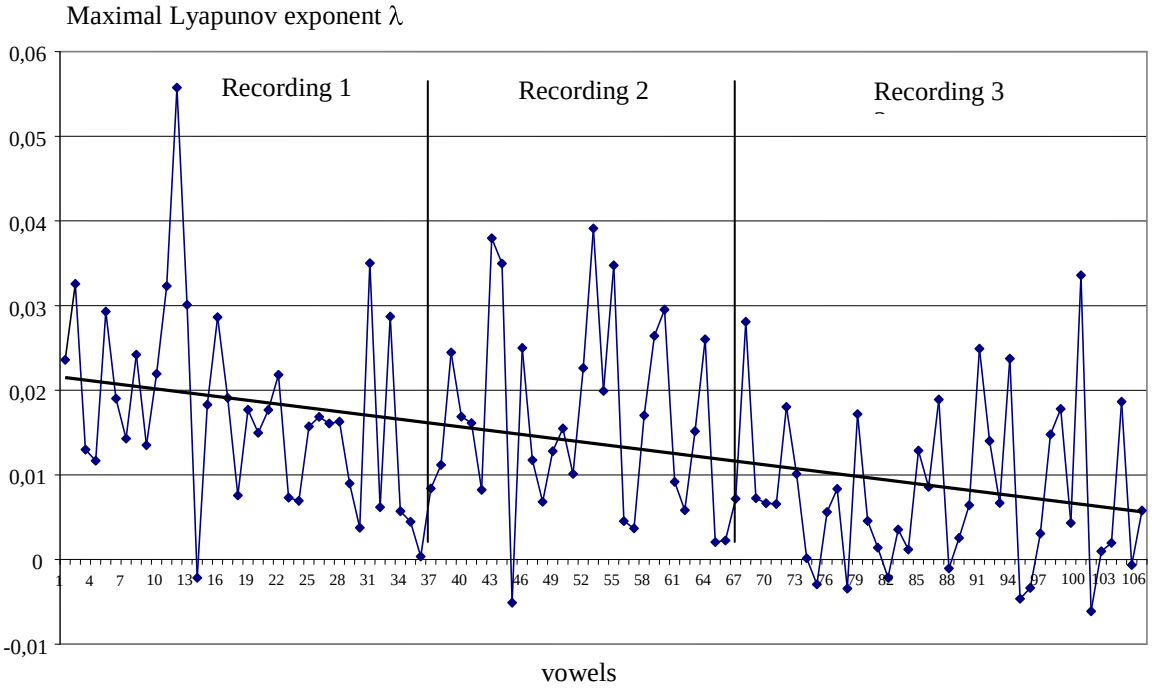


Figure 3.2.3.3. Évolution du plus grand exposant de Lyapunov au cours des trois phases de l'expérimentation sur l'inertie du sommeil et droîte de régression [5, 33].

Parmi les paramètres spectraux étudiés, seul le centre de gravité spectral présente des variations notables. Après le réveil du pilote il décroit significativement de 32% en moyenne pour tous les types de voyelles segmentées.

L'expérience sur l'inertie du sommeil est celle qui a induit le plus de modifications vocales. C'est parce qu'elle mélange deux conditions favorables : la voix après un réveil soudain tout en effectuant des exercices psychomoteurs. L'analyse de la voix peut donc prendre part aux études sur le comportement des pilotes après un endormissement lors des vols long-courriers.

3.2.4. Fatigue Physique (co-encadrement D.E.A M.Rolland 1993);

Elle a été provoquée à l'aide d'un vélo d'appartement selon un protocole d'effort évitant un essoufflement excessif du locuteur et avec une mesure simultanée du rythme cardiaque.

Les sujets de l'expérience pédalent jusqu'à ce que leur pouls atteigne la fréquence cardiaque maximale à savoir (220 moins l'âge) à plus ou moins 10%. Pour atteindre cette fréquence, la charge physique est augmentée régulièrement : les cyclistes doivent pédaler de plus en plus vite et les forces de frottement du pédalage sont régulièrement accrues.

Depuis l'état de repos, les enregistrements de la voix sont effectués d'abord toutes les vingt pulsations environ, après que le locuteur ait repris son souffle, puis toutes les dix pulsations environ jusqu'à la fréquence cardiaque maximale. Le cycliste doit prononcer les cinq premières voyelles orales de la langue française. Il s'arrête pour cela de pédaler à chaque enregistrement.

Les résultats montrent une augmentation de la fréquence fondamentale moyenne avec le rythme cardiaque pour l'ensemble des cinq voyelles et pour les sept locuteurs testés. La figure 3.2.4.1 en est une illustration pour deux des sept sujets.

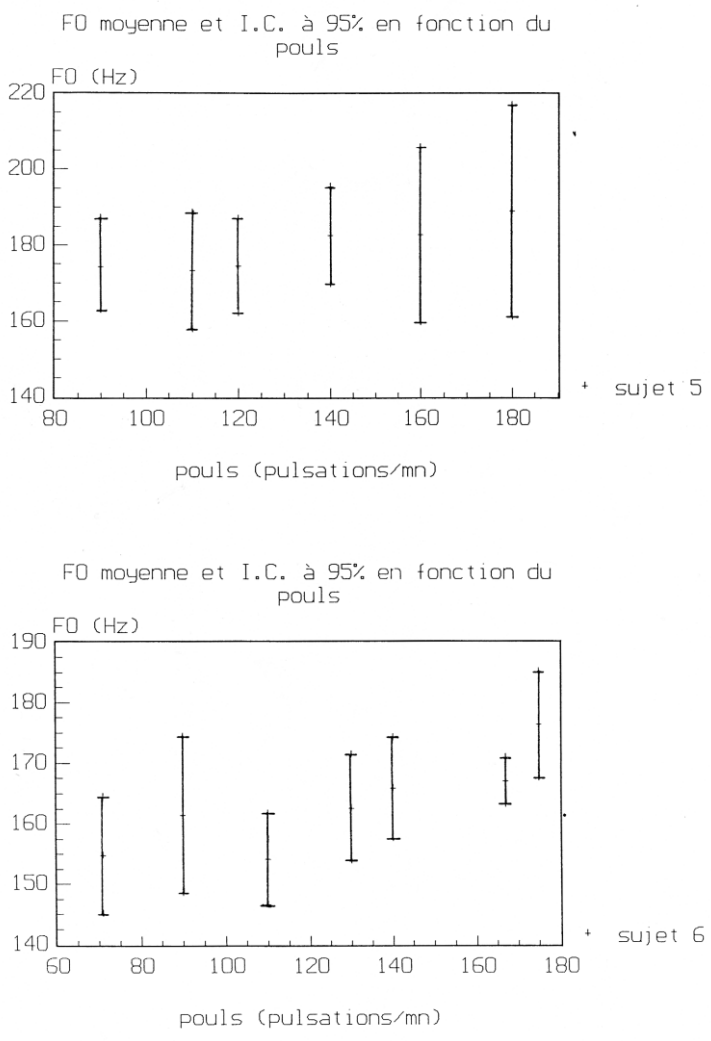


Figure 3.2.4.1 : Évolution de la fréquence fondamentale moyenne et de l'intervelle de confiance à 95% en fonction du rythme cardiaque pour deux des sept sujets testés (valeurs moyennes et intervalles de confiance à 95%) (D.E.A M.Rolland).

L'analyse des évolutions relatives des deux grandeurs par rapport à l'état de repos révèle des disparités entre les locuteurs. Une corrélation linéaire n'est réaliste que pour trois sujets sur sept (figure 3.2.4.2).

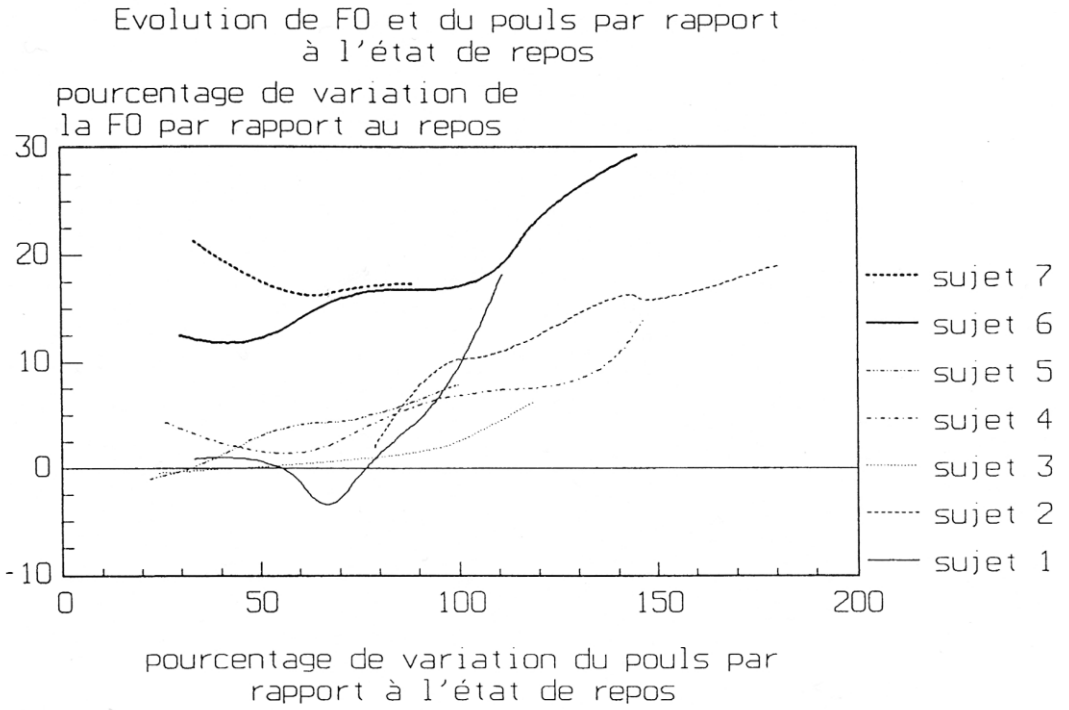


Figure 3.2.4.2 : Évolution conjointe de la fréquence fondamentale moyenne et du rythme cardiaque exprimés en pourcentage de variation par rapport à leur valeur respective au repos pour les sept sujets de l'expérimentation (D.E.A M.Rolland).

La conclusion de l'étude est que la fréquence fondamentale augmente avec le rythme cardiaque sans que les sujets parlent corrolairement avec une plus grande force sonore relative (un contrôle du niveau a été effectué).

L'étude ne prétend pas à conclure que toutes les formes de fatigue physique provoquent des modifications vocales. Il en est de même pour l'interprétation des résultats par rapport au rythme cardiaque, lequel est propice à varier sous l'effet de beaucoup d'autres facteurs.

3.3. Domaine de la Psychologie

Deux études ont été entreprises hors du contexte aéronautique. Elles appartiennent au domaine de la psychologie. La première observe les effets vocaux du test de Stroop. La seconde s'attache à estimer si la prononciation à la vue d'images peut contenir des signes non linguistiques de l'émotion qu'elles suscitent.

3.3.1. Test de Stroop (partenaires : laboratoires de Phonétique de l'Université de Mons-Hainaut (Belgique) et de l'Université de Barcelone (Espagne)) 1994 à 1996 [4], [15], [16]); co-encadrement de la thèse de doctorat de B.Gramatica, 1996.

Ce test, nommé également test Mot-Couleur, provoque un effet d'interférence de la lecture d'un mot sur la dénomination de sa couleur. Par exemple lorsqu'il s'agit de dénommer la couleur du mot « rouge » alors qu'il est écrit en vert : le mot « rouge » est écrit à l'encre verte et le locuteur doit prononcer « vert » (Test de stress, 1975 dans §5.2.5 page 142). L'objectif des analyses acoustiques est de rechercher des modifications du signal par rapport à une prononciation sans interférence.

Le test est divisé en trois phases. Pour les deux premières, le sujet prononce respectivement les noms de couleur qu'il lit puis les couleurs qu'il voit dessinées à l'intérieur de carrés. La troisième phase consiste à prononcer la couleur du mot écrit alors que celui-ci désigne une couleur différente. Ce test est considéré comme un test de stress par les psychologues car il provoque un effet cognitif entre la tâche de lecture plus naturelle et la tâche d'énonciation.

Les résultats de l'analyse de voyelles segmentées dans les couleurs prononcées montrent une augmentation significative de la fréquence fondamentale pour la troisième phase (figure 3.3.1.1) : 164,8 Hz en phase 3 contre 127,0 Hz et 130,5 Hz pour les phases 1 et 2.

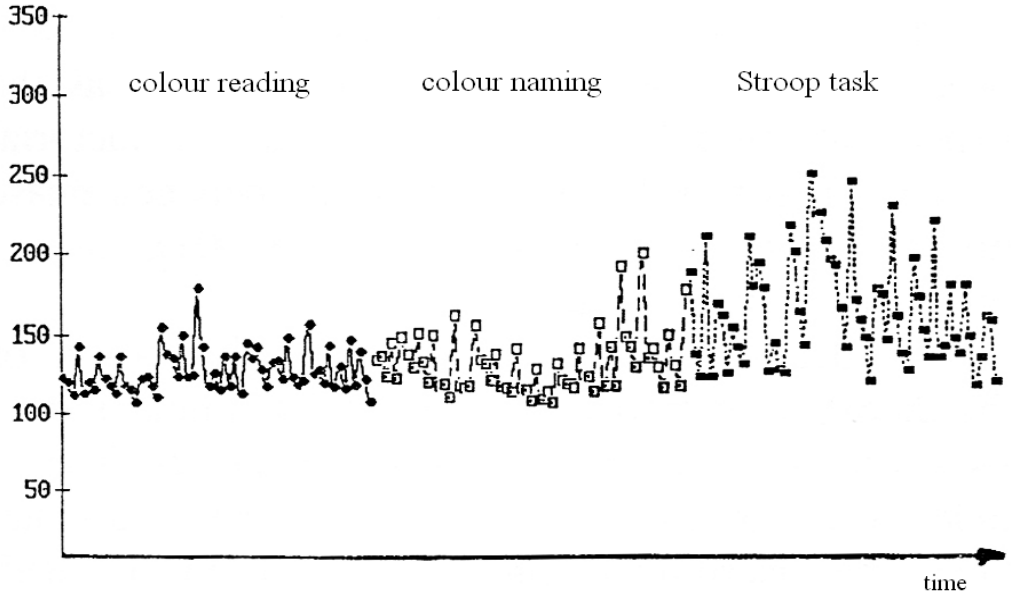


Figure 3.3.1.1 : Fréquence fondamentale moyenne (Hz) pour les trois phases du test de Stroop [4].

La dispersion de la fréquence fondamentale moyenne est également accrue : $\sigma = 8.9$ Hz pour la phase de lecture, $\sigma = 12.4$ Hz pour celle de reconnaissance et $\sigma = 21.7$ Hz pour la tâche complexe. De même, l'écart-type par phonème, c'est-à-dire celui de la distribution de la fréquence fondamentale par tranche d'analyse passe de 3.3 Hz à 11.2 Hz en moyenne soit une variation de 239%. Le jitter est également modifié dans le même sens avec une variation de plus de 100% (thèse B.Gramatica, 1996).

Un nouveau paramètre est introduit par l'équipe belge : $\mu = 2.f_c / (f_i + f_f)$ où f_i f_c et f_f sont respectivement les valeurs de la fréquence fondamentale pour les parties initiales, centrales et finales du son voisé [4]. Il a été constaté que l'indice μ devenait supérieur à 1 en phase 3 ($\mu = 1,3$) montrant ainsi une valeur centrale plus élevée qu'aux frontières, alors que pour les phases précédentes il restait voisin de 1 (1,07 en phase 1 et 1,3 en phase 2).

Des modifications spectrales ont pu également être mesurées. S'agissant de la fréquence des trois premiers formants, les sens de variation diffèrent selon la voyelle considérée.

La fréquence d'équilibre spectral (§ 3.5.2 pour sa définition page 87) de la moitié des voyelles prononcées en phase 3 a des valeurs situées en moyenne fréquence contrairement aux autres qui possèdent des fréquences d'équilibre spectral situées aux bornes inférieures ou supérieures du spectre [4]. Enfin, l'aire minimale mesurée à la fréquence d'équilibre spectral diminue avec la complexification de la tâche à effectuer (thèse B.Gramatica, 1996).

Le centre de gravité spectral augmente entre les phases initiales et la phase 3.

Mais si ces deux derniers critères sont très sensibles à la condition expérimentale « stress » de la phase 3, ils le sont également au type de voyelle prononcée.

Le test de Stroop a servi de support pour l'analyse des consonnes. Les paramètres mesurées ne permettent pas de distinguer la phase trois du test des deux précédentes (durée d'établissement du voisement, concentrations fréquentielles d'énergie dans la barre d'explosion) (thèse B.Gramatica, 1996).

La voix est donc modifiée par le test de Stroop tant à la fois dans le plan dynamique que dans le plan spectral des voyelles. Cette conclusion corrobore celle des études antérieures (§2.2.3 page 39). Des différences apparaissent sur le sens de variation de la fréquence des formants mais la fréquence fondamentale augmente dans toutes les expériences.

3.3.2. Banque d'images (partenaire : R. Da Silva Neves, enseignant-chercheur en Psychologie Cognitive et soutien du Conseil Scientifique de l'Université Toulouse 2 dans le cadre des Programmes Ponctuels de Recherche P.P.R) 2000 à 2002 [18], [19], [20] ;

Cette banque d'images (International Affective Picture System I.A.P.S) mise au point par le « Center for Emotion and Attention » de l'Université de Floride présente l'avantage de classer et de noter les images selon le degré émotionnel qu'elles induisent. L'expérimentation de laboratoire a conduit les spectateurs à parler en regardant les images selon un discours contrôlé afin de tenter d'établir un lien entre les notes respectives et les valeurs des grandeurs acoustiques mesurées.

Après l'apparition d'un point de fixation sur l'écran d'ordinateur, chaque image est examinée par le sujet pendant six secondes puis une phrase apparaît en surimpression pendant quatre secondes. Le sujet doit la prononcer après qu'un signal sonore (« bip ») ait été diffusé (quatre secondes après la disparition de la phrase à l'écran).

La phrase commence toujours par une structure Consonne1–Voyelle–Consonne2 (C1VC2). Cette structure désigne toujours un élément de l'image.

S'il s'agit de la photographie d'un visage mutilé, la phrase est « Kut est ensanglanté ».

Les structures CVC sont : [tip], [pit], [pak], [kap], [tyk], [kyt]. Les voyelles sont choisies pour couvrir un large domaine fréquentiel. Le choix des consonnes repose sur leur aptitude à être de nature impulsive afin de faciliter les repérages temporels qui vont suivre.

Les images proviennent de la banque d'image I.A.P.S (P.J.Lang, 1997 dans §5.2.2 page 126). Chacune d'entre elles possède deux notes comprises entre 0 et 10 qui traduisent leur impact émotionnel en termes de valence affective du stimulus (« pleasure » : P) et d'excitation induite par ce stimulus (« arousal » : A).

24 images ont été sélectionnées mais seulement 12 sont exploitées. Ce sont six images qui possèdent des notes élevées dans les deux dimensions précédentes notées A+/P+ et six images appartenant au groupe A+/P- qui ont une note P faible et une note A élevée.

Les images A+/P+ sont à caractère érotique, et les images A+/P- présentent par exemple des scènes d'accidents avec des blessures corporelles.

Les structures C1VC2 sont prononcées après chaque image A+/P+ et A+/P-.

Le déroulement de l'expérience est entièrement automatisé. Le sujet gère lui-même le passage d'une image à la suivante en cliquant simplement sur la souris. Une phase d'entraînement est prévue pour que le locuteur se familiarise avec le procédé.

En préalable, deux tests psychologiques sont effectués par les participants pour estimer leur état d'anxiété. Le locuteur dont la voix est analysée est retenu parce que les psychologues ont observé qu'il présentait une agitation comportementale attestant du trouble provoqué par les images.

Sur les structures CVC qui introduisent chaque phrase, deux durées sont calculées : d(C1V) et d(C1C2). Ce sont respectivement la durée entre l'amplitude maximale de C1 et celle de V et la durée entre l'amplitude maximale de C1 et celle de C2.

Deux mesures d'énergie sont effectuées. La première est celle du rapport noté E_{av}/E_{ap} entre l'énergie contenue dans le segment temporel d(C1V) et l'énergie contenue dans celui d(C2V). La seconde est appelée temps d'équilibre énergétique (TEE). C'est le temps qui sépare le segment d(C1C2) en deux parties d'égale énergie. Cette mesure ne prend de sens qu'en proportion de la durée d(C1C2). C'est pourquoi le rapport TEE/d(C1C2) est calculé.

Lorsque ces diverses grandeurs sont comparées entre les prononciations neutres, c'est-à-dire sans l'apparition des images, et celles après leur diffusion, on note que les mesures permettent d'isoler les deux corpus (figures 3.3.2.1 et 3.3.2.2).

Le résultat n'est pas aussi net pour les durées d(C1V) et d(C2V) en fonction des deux autres paramètres énergétiques. Mais l'examen des durées d(C1V) en fonction du type d'image révèle qu'elles varient du simple au double entre les images à la note élevée en termes de plaisir (A+/P+) et les images à la note faible (A+/P-).

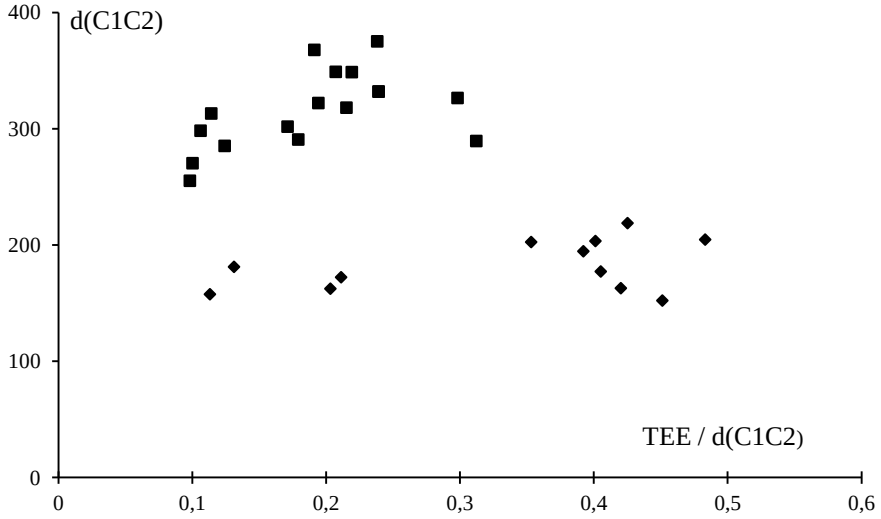


Figure 3.3.2.1 : Représentation des valeurs de d(C1C2) en fonction du paramètre TEE/d(C1C2). Les carrés correspondent aux prononciations « au repos » et les losanges à celles après la vue des images [20].

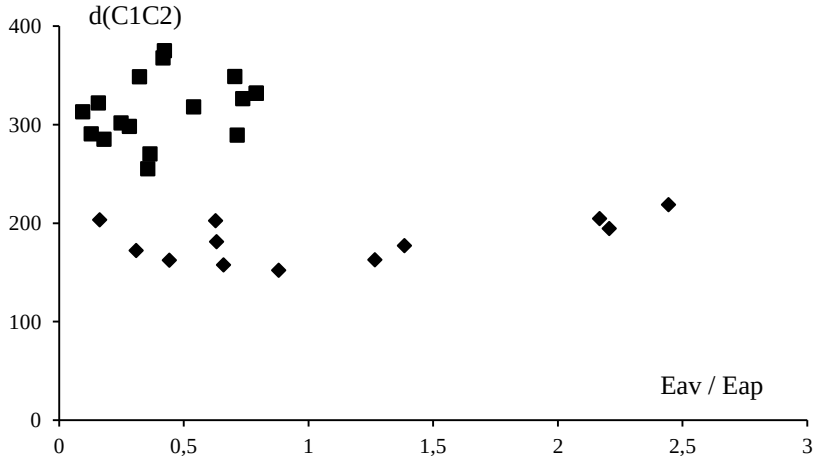


Figure 3.3.2.2 : Représentation des valeurs de d(C1C2) en fonction du paramètre E_{av}/E_{ap}. Les carrés correspondent aux prononciations « au repos » et les losanges à celles après la vue des images [20].

Avec ce même sujet et six autres, l'expérimentation a été conduite en s'attachant à mesurer la fréquence fondamentale moyenne, sa dispersion et le jitter, pour la voyelle de chaque structure CVC. Les images diffusées appartiennent toutes au corpus A+/P+.

Aucune variation significative n'a pu être observée entre les prononciations au repos et après ces images. Il est probable que la procédure expérimentale crée un effet de masque sur la structure sonore : la part émotionnelle de l'expression vocale peut avoir disparu avant que le « bip » ne donne l'indication sonore pour s'exprimer. En revanche, des variations de durée des structures lexicales apparaissent.

A l'instar de ce qu'est le réflexe audio-phonatoire, les premiers éléments de la possible existence d'un réflexe visio-phonatoire en lien avec l'émotion sont posés.

Les deux expériences mises en place dans le domaine de la psychologie apportent des résultats probants et viennent renforcer l'affirmation que le signal vocal est à même d'être modifié par toute sorte de phénomènes.

Il peut donc être apte à en détecter la présence mais cette sensibilité générale rend difficile l'analyse des causes, surtout lorsque les agents perturbateurs de l'état du locuteur peuvent se superposer ou se combiner.

3.4. Synthèse des Travaux Expérimentaux

Elle ne réside pas dans l'exposé de leur capacité à généraliser les résultats obtenus puisque les expérimentations ont le plus souvent été conduites pour un locuteur unique voire quelques uns. Il n'y a donc pas matière à conclure que tel ou tel paramètre acoustique possède l'aptitude à détecter tel ou tel type de perturbation psychophysiologique en établissant par exemple des seuils de variation au delà desquels la modification du signal vocal rend compte d'un trouble individuel.

En revanche, la diversité des situations expérimentales examinées et conduites avec des locuteurs différents peut attester ou pas de l'existence de modifications vocales dues à ces divers facteurs perturbateurs, et dans l'affirmative, ouvrir les pistes de recherche vers les propriétés du signal et les paramètres acoustiques les plus pertinents.

La synthèse des travaux expérimentaux montre que la voix est modifiée par le contexte psychophysiologique dans lequel se trouve le locuteur et en particulier dans les situations aéronautiques qui forment la majorité des expériences étudiées. Cette première conclusion corrobore les résultats des autres recherches dans le domaine. Elle les renforce et y apporte une contribution par sa spécificité aéronautique, généralisable au secteur plus vaste des transports.

Le deuxième élément de synthèse a trait à la méthodologie expérimentale et d'analyse des données.

Pour statuer sur le caractère singulier d'une prononciation, un ou des éléments de comparaison sont nécessaires. La référence vocale du locuteur doit être établie en dehors de tout facteur extérieur perturbateur de son état général. On parle de période « neutre » ou de « repos ».

Les expérimentations ont montré qu'il est préférable de la considérer ainsi, plutôt qu'immédiatement en amont ou en aval de la situation étudiée. Le contexte expérimental influe en effet lui-même sur la qualité de la voix, indépendamment de toute considération acoustique de prise de son. Ce fut le cas pour les expériences en simulateur de vol pour lesquelles la fréquence fondamentale moyenne dans le poste de pilotage était déjà différente et supérieure à celle mesurée avant d'y pénétrer (de même pour l'expérience de conduite automobile). Ces remarques ne signifient pas qu'il faille pour autant se priver d'une base de données dans le laps de temps qui précède immédiatement l'expérience.

Pour l'analyse des données, les comparaisons statistiques de moyennes d'un paramètre entre différentes phases, y compris celle au repos, ne s'avèrent pas pertinentes pour traduire

au mieux d'éventuelles variations anormales. Elles semblent l'être lorsque l'effet sur la voix est important mais elles sont très dépendantes du choix de la longueur des structures vocales à moyenner. Si modification vocale il y a, sa durée est inconnue. Se pose alors le problème de déterminer la frontière temporelle au delà de laquelle tout effet vocal a disparu, surtout lorsque le discours est naturel et non contrôlé par l'expérience.

Une observation continue des variations des paramètres acoustiques s'est imposée au cours des recherches. Elle n'obère pas les comparaisons de moyenne mais elle ne gomme pas de brèves modifications. L'expérience a montré que les expressions vocales perturbées peuvent être de courte durée, comme en témoigne l'analyse de l'enregistreur phonique, sauf lorsqu'elles sont entretenues par l'expérience elle-même, comme pour le test de Stroop qui d'ailleurs en fixe aussi l'instant d'apparition.

Ce choix de méthode d'analyse des données est d'ailleurs conforté par le fait que l'opération de segmentation des voyelles conduit à ne traiter qu'une information partielle du signal total. Il est imposé par les analyses spectrales réalisées en parallèle, mais pourrait ne pas être nécessaire pour certains paramètres du plan dynamique comme la fréquence fondamentale pour laquelle un suivi plus fin est possible.

Un dernier élément plaide en la faveur d'un suivi plutôt que d'un moyennage. Il s'agit de la dispersion associée aux valeurs moyennes mesurées. Très souvent en effet, les calculs de moyennes réalisés ont conduit à observer des dispersions élevées traduisant ainsi le caractère très fluctuant des paramètres acoustiques dans les situations perturbantes étudiées.

Le troisième et dernier élément de cette synthèse des travaux traite des paramètres acoustiques. C'est le plan dynamique du signal qui possède les caractéristiques les plus sensibles aux phénomènes étudiés. Dans le plan spectral et des phases les modifications ne sont pas systématiques et présentent souvent des sens de variation différents suivant le type de facteur perturbateur.

La fréquence fondamentale apparaît comme le paramètre acoustique qui varie systématiquement sous l'effet du contexte. Sa valeur moyenne par voyelle augmente. Mais comme il a pu être noté pour l'enregistreur phonique et l'étude de la charge de travail, un autre contexte provoque le même type d'augmentation en amplitude et en sens. Il est lié à l'intention du locuteur vis à vis de son interlocuteur : voix interrogative, voix pour convaincre par exemple.

Lorsque les deux contextes sont présents, les mesures de fréquence fondamentale ne permettent pas de séparer leur influence respective. L'observation conjointe d'autres types de paramètres (jitter, spectraux …) ne permet pas non plus de résoudre le problème pour l'instant. On peut néanmoins considérer que l'intonation participe à l'expression sonore de l'émotion.

La proposition de seuils relatifs de variation est possible, au delà desquels elle peut être considérée comme inhabituelle, mais elle n'a de sens qu'avec une extension du nombre de locuteurs testés.

Le plan dynamique offrant le plus grand potentiel de sensibilité, les travaux se sont orientés vers la proposition de nouveaux indicateurs inscrits dans ce plan. Il s'agit du D.A.L avec ses paramètres associés et de la fréquence de modulation des maxima d'amplitude (cf § 3.5.3 page 95 et §3.5.4 page 101 pour leur définition). Leur mesure sur les données expérimentales a donné quelques résultats prometteurs.

Les paramètres de la micro prosodie, comme le jitter et la dispersion des valeurs locales de la fréquence fondamentale F_0, ont été eux aussi sensibles mais dans une moindre mesure que F_0. Il en est de même pour le plus grand exposant de Lyapunov.

D'un point de vue spectral, il n'est pas possible de conclure aussi positivement malgré le nombre important de grandeurs examinées.

L'analyse souffre sans doute de sa dépendance au type de voyelle diminuant la quantité d'information pour le traitement : les comparaisons spectrales sont réalisées pour les mêmes voyelles. Les grandeurs classiques liées aux formants (fréquences, amplitudes, largeurs de bande) n'ont que très peu varié.

En revanche, comme pour la fréquence fondamentale et son contour temporel, il semble que la forme de l'enveloppe spectrale puisse être mise en relation avec les perturbations. C'est pourquoi un nouvel indicateur a été proposé autour de la distribution cumulée de niveaux sonores dans le spectre (§ 3.5.2 page 87 pour sa définition).

Les indicateurs plus classiques de la forme spectrale que sont le skewness et le kurtosis, mais aussi les grandeurs qui en dépendent comme la pente, le centre de gravité spectral, la fréquence d'équilibre énergétique et divers rapports d'énergie ont été mesurés. Les conclusions sont très variables d'une expérience à l'autre sans pour autant que ces paramètres montrent des variations remarquables.

Les résultats permettent de conclure que la forme spectrale subit des altérations mais sans qu'aucun paramètre n'en rende systématiquement compte.

Les travaux expérimentaux sont majoritairement inscrits dans le secteur aéronautique. L'exercice du pilotage peut provoquer fatigue, charge de travail, stress, émotion, perte de vigilance. Ces perturbations ont des spécificités propres au métier : fatigue des vols court-courriers, stress et émotion lors des dysfonctionnements, inertie du sommeil. Elles ont été expérimentées et il ressort que l'analyse de la voix peut être sérieusement envisagée pour non seulement participer à l'étude de ces phénomènes mais aussi envisager d'en détecter l'apparition voire d'en estimer le degré.

3.5. Paramètres Acoustiques Mesurés

3.5.1. Paramètres Classiques

Les caractéristiques acoustiques classiques du signal de voyelle n'ont pas toutes été mesurées simultanément pour toutes les expérimentations pour cause de progression dans la recherche et dans les moyens d'analyse et de calcul. Elles font ici l'objet d'une description de leur méthode de mesure avant de détailler dans les paragraphes qui suivent les trois indicateurs originaux proposés.

Tous ces paramètres sont aujourd'hui automatiquement mesurés par des programmes informatiques développés dans l'environnement Matlab® et aptes à traiter des listes de signaux de voyelles segmentées.

Les paragraphes qui suivent en présentent les conditions de mesure.

Segmentation, fréquence fondamentale et paramètres dérivés

La méthode de mesure de la fréquence fondamentale F_0 est fondée sur la détection des maxima d'amplitude du signal de la voyelle avec une fenêtre temporelle glissante. La durée de chaque cycle vibratoire résulte de l'estimation de l'intervalle de temps entre les pics successifs. On en déduit la fréquence fondamentale du cycle.

La fréquence fondamentale moyenne de la voyelle résulte de la moyenne des fréquences de chaque cycle. Cette valeur est comparée à celle obtenue par analyse cepstrale sur la durée totale du signal de voyelle et n'est validée que si l'écart est inférieur à 10%. Cette valeur de 10% est arbitrairement choisie suite aux essais de validation de l'algorithme de calcul. Elle représente un écart d'une dizaine d'Hertz.

La validation est acquise pour l'extrême majorité des signaux. Lorsque ce n'est pas le cas, un examen détaillé du signal est effectué pour en déterminer les causes et corriger la mesure.

Les cycles vibratoires retenus pour le calcul de F_0 sont ceux du segment de voyelle extrait.

La segmentation est effectuée à l'aide d'un logiciel d'édition audio (Audacity®) et ne conserve que la partie quasi stationnaire du signal. L'attaque, l'extinction, lorsqu'elles existent, et la transition avec les phonèmes adjacents sont éliminées car elles présentent une plus grande variabilité naturelle.

Lorsque le débit phonatoire est élevé, le plus souvent en discours naturel, l'opération de segmentation n'est pas toujours possible car les durées des voyelles peuvent être très courtes.

En règle générale, la voyelle n'est conservée pour l'analyse que si la segmentation permet de disposer d'au moins quatre périodes consécutives.

La méthode de détection de F_0 peut être qualifiée de robuste et de fiable suite aux nombreuses vérifications directes sur le signal temporel.

De cette détection des pics d'amplitude sont déduites les mesures de jitter, de shimmer, d'écart type de la distribution des valeurs de F_0, de coefficient de variation.

Micros – tremblements des cordes vocales

Leur mise en évidence a été réalisée par un instrument spécifique analogique, le P.S.E (« Psychological Stress Evaluator ») qui par la suite a été adapté et commercialisé dans une version numérique pour la détection de mensonge, le C.V.S.A (« Computer Voice Stress Analyser »).

Une version du C.V.S.A, développée dans l'environnement Matlab®, a été proposée (J.H.L.Hansen, 2000, §5.2.1 page 124). Elle est reprise ici et adaptée pour être compatible avec la fréquence d'échantillonnage utilisée et pour s'inscrire dans les algorithmes développés.

Son principe général est de simuler le ralentissement de la vitesse de défilement de la bande d'enregistrement par un suréchantillonnage sans correction de temps. Le facteur de suréchantillonnage est inversement proportionnel au facteur de ralentissement.

Après quoi, le signal, dont ne sont conservées que les alternances positives, est filtré passe-bas avec un filtre dont la réponse en fréquence reprend les caractéristiques préconisées par le P.S.E et le C.V.S.A : fréquence de coupure égale à 12 Hz et fréquence de réjection égale à 15 Hz.

Les essais effectués permettent de retrouver les allures des diagrammes proposés par les appareillages commercialisés. C'est ce qu'illustre la figure 3.5.1.1 où une fluctuation de la composante continue, indicatrice de l'absence de stress, est observée.

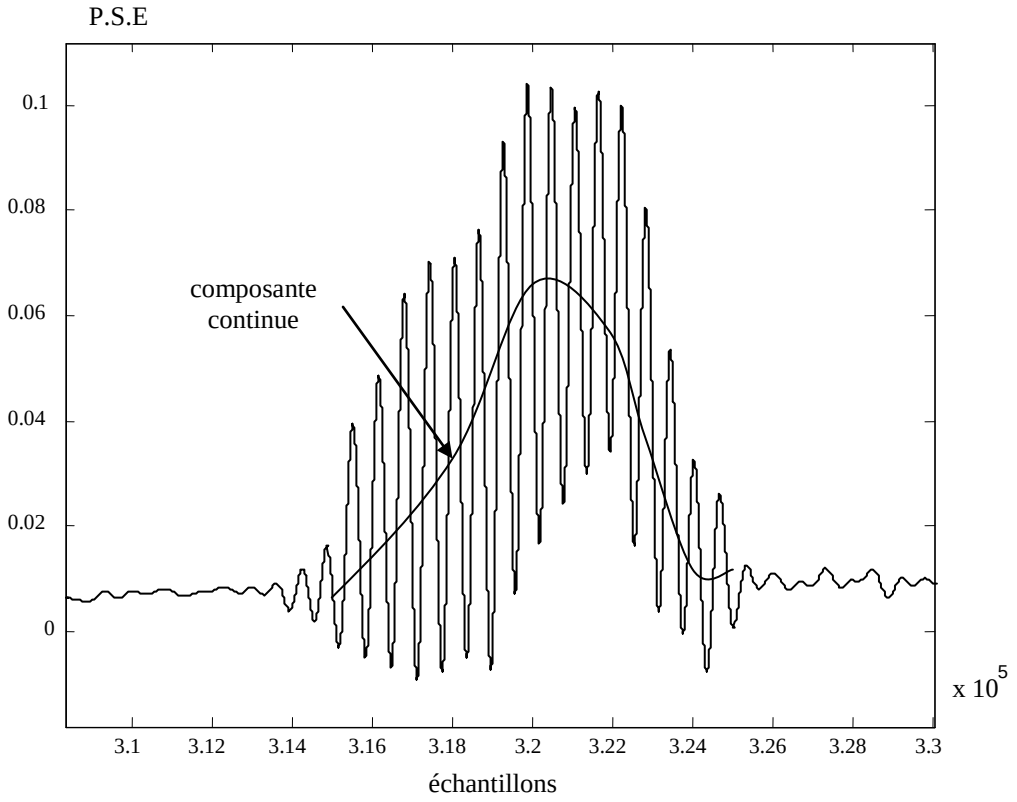


Figure 3.5.1.1 : Exemple de diagramme P.S.E obtenu par l'algorithme de calcul mis en place. Cas d'une voyelle prononcée dans un état de neutralité émotionnelle.

C'est à partir de ces diagrammes que l'expérimentateur effectue un examen visuel pour statuer sur la présence ou pas de stress dans le discours (P.S.E). Dans le cas du C.V.S.A et d'appareillages similaires, des calculs sont effectués pour attribuer un score à la forme obtenue. Ce score traduit les plus ou moins grandes irrégularités de la composante continue du signal démodulé en fréquence.

Lors de la programmation, seule la visualisation des diagrammes a été mise en place mais diverses conditions de mesure ont été testées. Il n'a pas encore été développé une méthode de notation quantitative des diagrammes comme cela a pu être proposé dans la littérature relative au P.S.E.

Exposant de Lyapunov

La méthode de calcul utilisée et programmée est celle de H.Kantz (H.Kantz 2005, §5.2.5 page 142) parce qu'elle est adaptée à la mesure du plus grand exposant de Lyapunov pour de courtes séries temporelles de données expérimentales. Le principe est le suivant.

Dans un système non chaotique, deux trajectoires initialement séparées d'une petite distance δx_0 seront séparées, au bout d'une certaine durée T, d'une distance δx proportionnelle à la distance initiale. Les trajectoires se séparent proportionnellement au temps.

Dans un système chaotique, la séparation des trajectoires est plus rapide :

$$\delta x = \delta x_0 \, . \, e^{t/T} \quad (\text{avec } T>0)$$

Si d_j désigne la distance euclidienne entre deux points j et j' de la trajectoire, l'expression de d_j est donc la suivante :

$$d_j = d_j(0).e^{\lambda.i.\Delta t} \qquad (1)$$

où Δt est le pas d'échantillonnage et où $\Delta n=(i.\Delta t)$ représente la durée (en nombre d'échantillons) à laquelle est évaluée la distance de séparation d_j par rapport à la séparation initiale $d_j(0)$ à t = 0. λ est le plus grand exposant de Lyapunov. Il traduit le caractère chaotique s'il est positif. S'il est négatif, les trajectoires se rapprochent exponentiellement vers un point fixe stable. S'il est nul, il traduit l'existence d'un cycle limite stable (signal sinusoïdal).

L'expression logarithmique de l'équation (1) devient :

$$\langle Ln(d_j)\rangle = \langle\lambda\rangle.\Delta n + \langle Ln(d_j(0))\rangle \qquad (2)$$

La pente du tracé de $<Ln(d_j)>$ en fonction de Δn donne l'exposant de Lyapunov.

λ a la dimension inverse du temps (ou du nombre d'échantillons). L'horizon de Lyapunov ($1/\lambda$) possède quant à lui une dimension temporelle. Avec Δn en nombre d'échantillons, il convient de multiplier λ par la fréquence d'échantillonnage pour obtenir le plus grand exposant de Lyapunov en s^{-1}.

À partir d'un point de la trajectoire, la procédure consiste à sélectionner les « voisins » de ce point. Ceux-ci possèdent une distance euclidienne par rapport au point initial qui est inférieure à une distance ε fixée. Sur l'attracteur et avec une dimension de prolongement égale à 3, cela revient à définir une sphère de diamètre ε à l'intérieur de laquelle se trouvent les points du voisinage. Pour chaque Δn, le calcul de la moyenne du logarithme des distances entre le point considéré et ses voisins est effectué. On obtient ainsi $<Ln(d_j)>$ pour tous les Δn et on déduit λ du tracé (pente de la régression linéaire).

La validation de l'algorithme a été effectuée à partir de séries de données issues de fonctions dont la valeur théorique de λ est connue (fonction logistique et de Hénon) pour des longueurs de séries équivalentes à celles des voyelles.

En pratique, le temps de calcul peut devenir très long et les voyelles les plus longues ont été segmentées à nouveau pour conserver une taille de fichier inférieure à 10ko avec une fréquence d'échantillonnage de 44.1 kHz, soit une durée d'environ 110 millisecondes. Elles ont parfois été analysées par tranches lorsqu'au moins deux tranches de 10ko pouvaient être segmentées. Cette situation s'est présentée rarement : au plus, une voyelle sur six était dans ce cas pour l'analyse de l'enregistreur phonique.

Toutes les données expérimentales sont par nature bruitées et l'estimation de λ est très sensible au niveau de bruit dans le signal. De manière à supprimer les points bruités du voisinage, une sphère minimale est définie sur la trajectoire reconstruite (A.Giovanni 1999 §5.2.5 page 142, et [5] §5.1 page 121). Les voisins appartiennent au volume entre les deux sphères : la sphère de bruit de diamètre δ et la sphère de diamètre ε ($\delta<\varepsilon$). Le choix de δ et de ε varie suivant chaque situation expérimentale. Il dépend de l'estimation des variations du rapport d'amplitude entre le signal « voix + bruit » et le signal « bruit ».

En conditions de prise de son de laboratoire (inertie du sommeil), aucune sphère de bruit n'a été utilisée compte tenu des grandes valeurs de rapport signal vocal / bruit.

En présence de bruit de fond, comme pour l'analyse de l'enregistreur phonique, le choix de δ est important.

Si δ est trop grand, de véritables voisins des points de la trajectoire peuvent être ignorés pour le calcul de λ et si δ est trop petit, des voisins qui n'appartiennent pas à la trajectoire du signal vocal peuvent être pris en compte.

ε est un pourcentage de l'amplitude maximale « pic à pic » du signal et δ un pourcentage du rapport d'amplitude bruit / (signal vocal + bruit). Pour les expériences de laboratoire ε a été fixé à 2% pour toutes les voyelles segmentées.

Avec des valeurs plus basses le nombre de voisins qui est calculé pour chaque Δn devient faible et ignore certaines parties de l'attracteur.

Pour des valeurs plus élevées les résultats ne sont pas sensiblement modifiés.

En présence d'un bruit de fond non maîtrisé, mais faible, où la prise de son est réalisée par l'expérimentateur (vols court-courriers, conduite automobile), δ a été fixé à 2% et ε à 4%. Enfin pour l'enregistreur phonique, où la présence de bruit est plus importante essentiellement en raison du non contrôle de la prise de son, les conditions ont conduit à choisir δ = 15% et ε = 25%.

Pour le traitement automatique des ensembles de voyelles segmentées et pour conserver une reproductibilité dans les conditions d'analyse, le choix de la valeur d'arrêt de Δn pour le calcul de la régression linéaire a été figé et est demeuré constant pour chaque expérimentation. L'opération de validation de l'algorithme a conduit à une estimation convenable du plus grand exposant de Lyapunov pour Δn variant de 1 à 5. Ce choix est maintenu pour les voyelles qui possèdent des durées équivalentes à celle des séries temporelles utilisés pour la validation.

Les recherches à propos de la présence de chaos dans le signal vocal ne concluent pas toutes à l'existence d'un exposant de Lyapunov positif, synonyme de divergence des trajectoires et donc de chaos.

Aussi, plutôt que de s'intéresser au signe de λ et donc à conclure sur la nature chaotique ou pas du signal, il a été envisagé de représenter et d'étudier ses variations chronologiques ou de comparer ses valeurs entre différentes phases expérimentales. On ne considère donc plus λ d'un point de vue dichotomique mais comme un paramètre continument variable du signal vocal.

La dimension de prolongement pour la construction du plan des phases et la représentation de la trajectoire est fixée à 3. Il semble établi qu'en choisissant trois dimensions, la structure géométrique de l'attracteur est convenable et qu'elle peut être visualisée (H.Herzel 1994, §5.2.5 page 142).

Enfin le retard τ correspondant au délai temporel de reconstruction de la trajectoire est choisi égal au premier zéro de la fonction d'autocorrélation du signal de manière à garantir une indépendance entre chaque échantillon s(n) et s(n + k.τ) (k entier naturel positif) (M.T.Rosenstein 1993, §5.2.5 page 142).

Analyse spectrale

L'enveloppe spectrale moyenne de chaque voyelle est obtenue à partir d'un modèle auto régressif de chaque tranche de signal par la technique de prédiction linéaire (méthode d'autocorrélation) (J.D.Markel 1976, §5.2.5 page 142).

Les paramètres de l'analyse pour la fréquence d'échantillonnage f_e de 44,1 KHz sont : ordre p = 48, coefficient de pré accentuation μ = 0.98, fenêtre de Hamming de 512 échantillons avec un recouvrement de 50% entre les fenêtres de pondération successives.

Pour des conditions de bruit de fond important, le coefficient de pré accentuation est annulé car il provoque un renforcement excessif des hautes fréquences dont le niveau est déjà augmenté par la présence du bruit de fond ([29, 30, 31], §5.1 page 121).

Les coefficients de prédiction a(i) (i=0 à p) de chaque tranche de signal sont moyennés rang par rang pour l'ensemble des tranches de la voyelle.

L'enveloppe spectrale moyenne résulte alors d'une transformée de Fourier rapide (F.F.T) sur cet ensemble de coefficients moyens (1024 points soit une résolution de 43 Hz sur le spectre 0 - f_e /2).

Les calculs des paramètres spectraux s'effectuent uniquement sur la largeur de bande 0 –5000 Hz qui possède toute l'énergie du signal de voyelle.

Il s'agit du centre de gravité spectral, de la fréquence qui partage le spectre en deux bandes d'égale énergie ou fréquence d'équilibre énergétique, de la pente spectrale obtenue par régression linéaire, puis de la fréquence (par calcul de la dérivée seconde du spectre du modèle auto-régressif), de l'amplitude et de la largeur de bande des trois premiers formants et des moments normalisés du spectre (skewness (moment d'ordre 3) et kurtosis (moment d'ordre 4) par rapport au centre de gravité spectral). Le skewness rend compte des différences entre la forme spectrale avant et après le centre de gravité spectral. Le kurtosis est une mesure de l'applatissement du spectre.

Des distances spectrales sont également mesurées. Ce sont les distances entre deux ensembles de coefficients cepstraux exprimés en Mels (M.F.C.C).

Le signal de voyelle est divisé en tranches successives de 512 échantillons avec un recouvrement de 50%. Pour chacune d'entre elles, la procédure de calcul est la suivante : pré accentuation (μ = 0.98), pondération temporelle de Hamming, F.F.T (512 points), filtrage Mel (32 filtres entre 133 et 5000 Hz), logarithme népérien du signal filtré, transformée en cosinus discrète (J.Benesti 2008, § 5.2.5 page 142).

Les 32 coefficients M.F.C.C ainsi obtenus pour chaque tranche sont moyennés pour chaque rang sur l'ensemble des tranches comme pour les coefficients de prédiction. On obtient ainsi un vecteur de 32 coefficients moyens associé à chaque voyelle.

La distance spectrale euclidienne peut alors être calculée entre deux ensembles de vecteurs. Par exemple le vecteur résultant de la moyenne de ceux des voyelles prononcées en situation de repos avec le vecteur résultant de la moyenne de ceux prononcés en état perturbé.

3.5.2. La Fréquence d'Équilibre Spectral (F.E.S)

Le calcul d'une fréquence qui partage le spectre en deux bandes ou qui en est le centre de gravité est communément appliqué pour l'analyse de spectres. Une seule valeur ne saurait caractériser le spectre mais sa comparaison dans différentes situations de prononciation peut en faire un facteur lié aux perturbations vocales.

L'objectif est de traduire d'éventuels déplacements d'énergie par la mesure d'une seule grandeur.

Pour tous ces types de mesure, le spectre est borné et les grandeurs sont indépendantes de la méthode d'analyse spectrale.

Deux approches sont possibles :

- l'approche « rapport d'énergie » qui consiste à fixer une fréquence dans le spectre (1000 Hz par exemple) et à étudier l'évolution du rapport d'énergie bande basse – bande haute fréquence ou bien, à rechercher la fréquence qui le partage en deux bandes d'égale énergie (fréquence d'équilibre énergétique). C'est cette dernière méthode qui a été retenue dans les expérimentations menées précédemment mais les deux répondent au même type de questionnement ;
- l'approche « centre de gravité » : au même sens qu'en mécanique, chaque fréquence est affectée d'un « poids » correspondant à son niveau sonore pour en rechercher le barycentre ou centre de gravité spectral.

La fréquence d'équilibre spectral (F.E.S) répond quant à elle à une troisième approche. Elle s'attache à prendre en compte la distribution cumulée des niveaux sonores dans deux bandes fréquentielles.

La F.E.S est la fréquence qui partage le spectre en deux bandes pour lesquelles les deux distributions sont les plus proches.

Il s'agit de distribution cumulée du niveau qui, parce qu'elle est toujours comprise entre 0 et 1, facilite la comparaison entre deux bandes, ce qui n'est pas le cas de la distribution simple. Elle permet donc la recherche de la fréquence pour laquelle les deux distributions cumulées sont les plus voisines.

Calcul de F.E.S

Considérons l'enveloppe d'un spectre comportant N fréquences discrètes entre f_{min} et f_{max}, l'histogramme du niveau sonore P(Li) peut être défini comme suit : $P(L_i) = (N_i/N)$ où N_i désigne le nombre de fréquences discrètes (ou raies) pour lesquelles le niveau sonore appartient à la classe de niveau de valeur centrale L_i.

$P(L_i)$ désigne donc la proportion de fréquences discrètes comprises dans la classe de niveau L_i et de largeur de classe ΔL. La probabilité cumulée, notée I_i, correspondant au niveau central L_i , est définie par la relation suivante :

Elle exprime le cumul des proportions des classes dont le niveau est inférieur ou égal à L_i.

Le diagramme représentant I_i en fonction de L_i est appelé Diagramme de Probabilité Spectrale Cumulée du niveau sonore (D.P.S.C), et I_i est nommé indice statistique.

La figure 3.5.1 est un exemple de calcul du diagramme. Le D.P.S.C résulte du lissage de l'histogramme cumulé par une régression polynomiale.

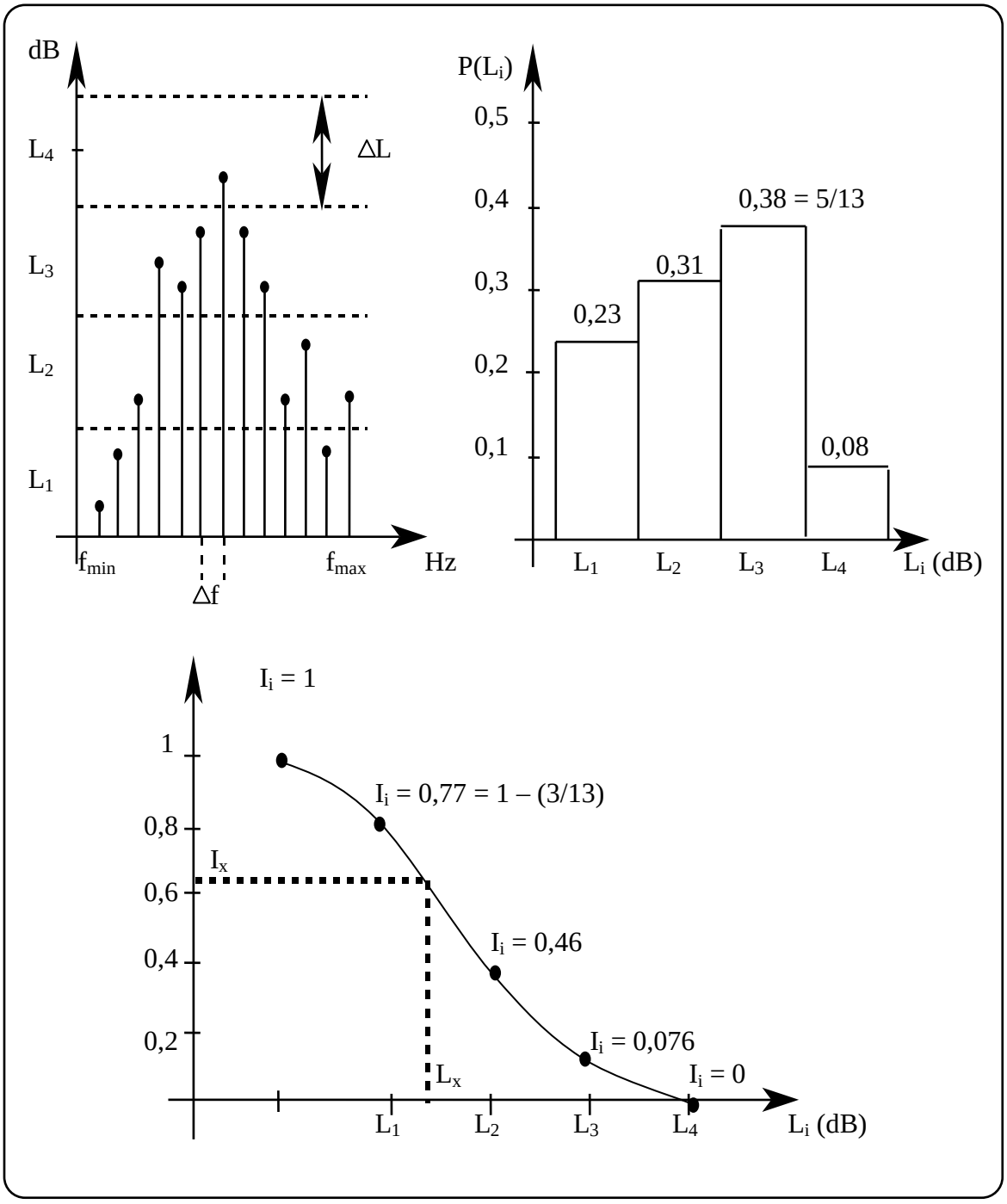


Figure 3.5.1: Exemple de calcul du Diagramme de Probabilité Spectrale Cumulée du niveau sonore à partir de l'enveloppe d'un spectre fictif de 13 raies [12].

En pratique, le D.P.S.C est calculé sur le spectre L.P.C moyen de chaque voyelle (§ 3.5.1 page 81). Mais le calcul peut être effectué à partir de n'importe quelle méthode d'analyse spectrale.

Pour obtenir la fréquence d'équilibre spectral, une fréquence du spectre est fixée pour calculer les deux D.P.S.C correspondants. L'aire qui les sépare est évaluée en procédant à la sommation de la différence en valeur absolue ΔI_i entre chaque valeur d'indice pour chaque valeur de niveau sonore L_i (figure 3.5.2). En parcourant le spectre fréquence par fréquence, la F.E.S est celle pour laquelle l'aire entre les deux D.P.S.C est minimale. Cela ne signifie pas pour autant que les distributions des niveaux sont identiques, elles sont au mieux ressemblantes.

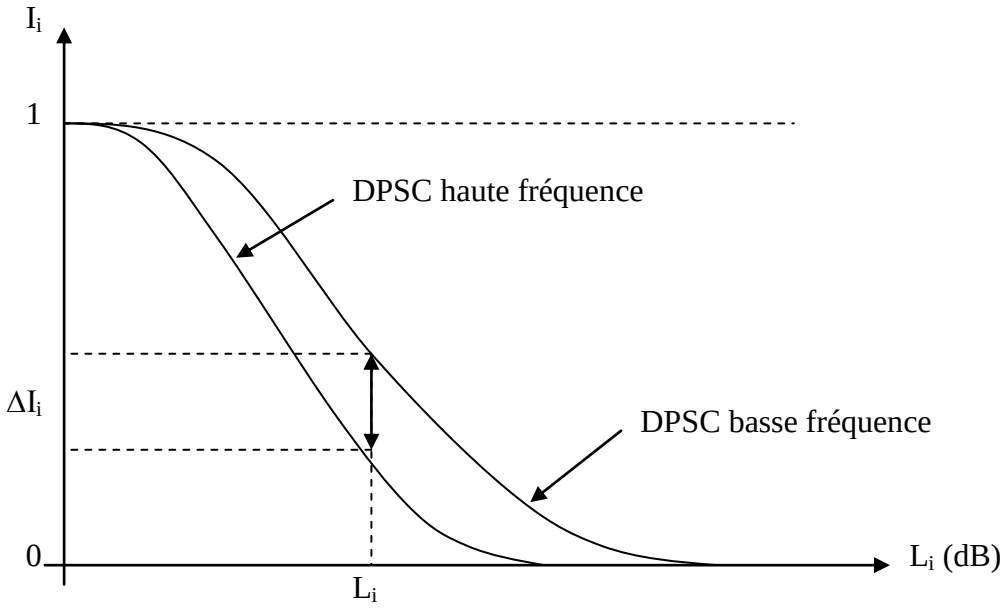


Figure 3.5.2 : calcul de l'aire entre deux D.P.S.C en sommant les ΔI_i pour chaque classe de niveau central L_i.

Propriétés

● L'histogramme du niveau sonore ne se prête pas à la comparaison de différentes voyelles ou groupes de voyelles. En effet, d'un phonème à l'autre les valeurs maximales et minimales du niveau peuvent être très différentes. Le D.P.S.C est quant à lui toujours borné entre 0 et 1 quelle que soit la dynamique du signal. Il est donc préférable de caractériser une voyelle par son D.P.S.C que par son histogramme.

● L'indice statistique I_i ($0 \leq I_i \leq 1$) du niveau sonore L_i représente le niveau atteint ou dépassé pour (I_i fois 100)% du domaine de fréquence. Par exemple si L_i = 20 dB correspond à l'indice I_i = 0,4 cela signifie que le niveau de 20 dB a été atteint ou dépassé pour 40% des fréquences du domaine fréquentiel que couvre le spectre.

Ainsi, I_i = 0 indique le niveau maximal du spectre et I_i = 1 le niveau minimal.

La pente du D.P.S.C renseigne sur la dynamique du signal. La détermination des niveaux correspondant aux indices statistiques possède une incertitude égale à la demi-largeur de classe.

● Le Diagramme de Probabilité Spectrale Cumulée permet d'estimer les largeurs de bande. L_i est le niveau dépassé pour (I_i.100)% du domaine de fréquence ($f_{max} - f_{min}$) (Hz). Cette propriété résulte de la définition même de la probabilité cumulée.

Dans l'exemple de la figure 3.5.1 le niveau L_2 est dépassé pour 46% de ($f_{max} - f_{min}$) (Hz). Le niveau L_4 n'est jamais dépassé (I = 0) et les classes de niveau inférieures à L_1 sont toujours dépassées (I = 1).

Ces considérations peuvent être traduites en termes de largeur de bande. Le niveau L_3 est dépassé pour 0,46.($f_{max} - f_{min}$). Autrement dit, la largeur de bande à –($L_4 - L_2$) (dB) vaut 0,46.($f_{max} - f_{min}$). (Hz).

Plus généralement, la largeur de bande B (en Hz) à –($L_{max} - L_x$) (dB) s'écrit: B = I_x.($f_{max} - f_{min}$), où L_{max} est la valeur centrale de la classe du niveau maximal atteint dans le domaine de fréquence étudié, soit de f_{min} à f_{max} .

Pour I_x = 0, c'est-à-dire pour $L_i = L_{max}$, on a B = 0 Hz, et pour I_x = 1, c'est-à-dire pour $L_x \le \left(L_{\min} - \frac{\Delta L}{2} \right)$, on a B = ($f_{max} - f_{min}$).

Enfin, pour L_x tel que ($L_{max} - L_x$) = 3 dB, B devient la largeur de bande classique à -3 dB.

Afin de vérifier la relation définissant la largeur de bande, l'enveloppe d'un spectre en forme de parabole a été simulée ; ceci dans le but d'appliquer la méthode à la mesure des largeurs de bande des formants. Son équation est de la forme $L(f) = af^2 + bf + c$, où L et f désignent respectivement le niveau et la fréquence et a,b,c sont des constantes.

La largeur de bande à -($L_{max} - L_x$) dB s'écrit alors (en Hz): $B = \left| \frac{\sqrt{b^2 - 4a(c - L_x)}}{a} \right|$

La comparaison entre la vraie valeur de B donnée par la relation précédente, et son estimation par la relation B = ($f_{max} - f_{min}$) est obtenue en calculant la valeur absolue de la différence de ces deux valeurs.

Il est nécessaire de procéder au tracé du D.P.S.C en ayant au préalable échantillonné en fréquence l'enveloppe du spectre L(f) (résolution Δf).

L'erreur d'estimation des largeurs de bande en fonction de la différence L_{max} - L_{min} pour une largeur de classe de niveau ΔL= 1 dB est inférieure à 1 Hz pour des résolutions fréquentielles Δf de 1Hz et de 0,1Hz. Elle augmente nettement si Δf est grand.

On note que l'erreur commise est d'autant plus petite que Δf et ΔL sont conjointement moins élevés. Dans ces conditions, elle peut devenir très faible (inférieure à 1 Hz) ([12] §5.1 page 121).

Sans précaution particulière, une largeur de bande déduite du diagramme peut ne pas correspondre à celle d'un formant, mais à la somme des largeurs de bande des résonances du domaine de fréquence.

Avant de calculer le D.P.S.C, il est donc nécessaire de limiter le domaine à l'intervalle de fréquence autour de la résonance. On en déduira alors toutes les largeurs de bande du formant.

Enfin, sous réserve que la résolution fréquentielle associée à l'échantillonnage de l'enveloppe spectrale, et la largeur des classes de niveau soient suffisamment faibles, les

erreurs d'estimation seront du même ordre de grandeur que celles indiquées par la simulation précédente.

● Une simulation d'enveloppes spectrales canoniques est effectuée pour mesurer simultanément la fréquence d'équilibre spectral (F.E.S), la fréquence d'équilibre énergétique (F.E.E) et le centre de gravité spectral (C.G.S). La première enveloppe est celle d'un spectre linéairement décroissant simulant une pente spectrale de voyelle. La seconde inclut dans la décroissance un pic de niveau. La troisième inclut trois pics de niveau et se rapproche ainsi de l'allure de l'enveloppe spectrale d'une voyelle.

Spectre linéairement décroissant :

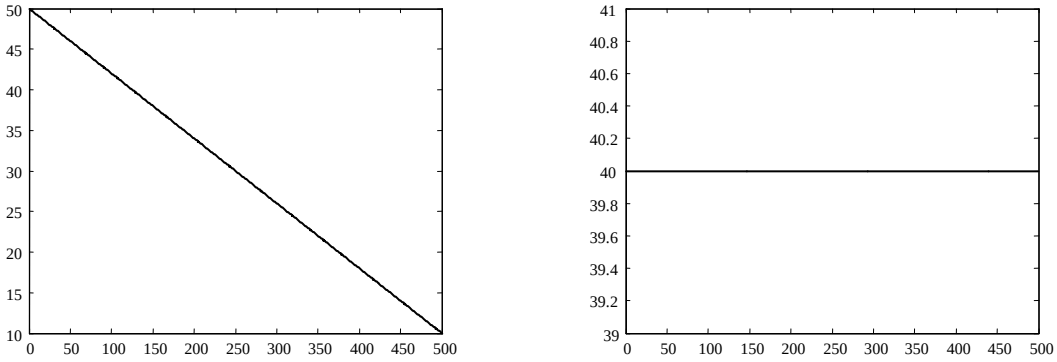


Figure 3.5.3 : Spectre (dB-Hz) linéairement décroissant (à gauche) et évolution de l'aire entre les D.P.S.C pour chaque fréquence (à droite) : F.E.S = 1 à 500 Hz, F.E.E = 38 Hz, C.G.S = 55 Hz.

L'exemple de la figure 3.5.3 peut être généralisé à toutes les formes spectrales de ce type qui conduisent donc à une F.E.S constante quelle que soit la fréquence. Les valeurs de F.E.E et C.G.S dépendent de la largeur de bande du spectre ce qui n'est pas le cas de F.E.S.

Spectre avec un pic de niveau :

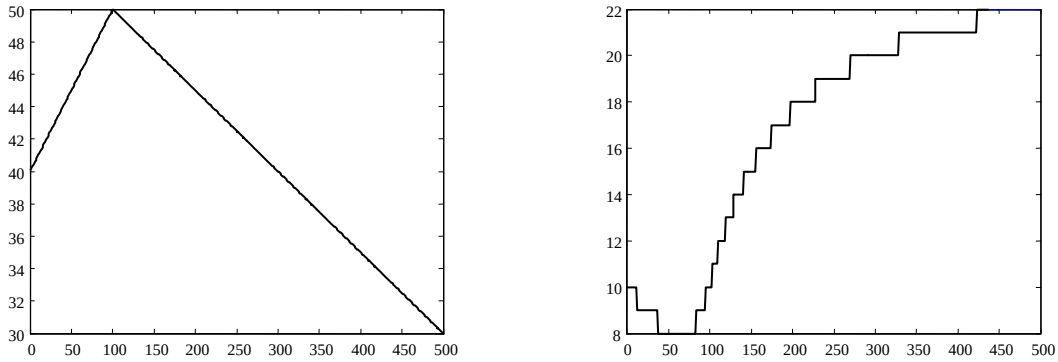


Figure 3.5.4 : Spectre (dB-Hz) avec un pic de niveau (à gauche) et évolution de l'aire entre les D.P.S.C pour chaque fréquence (à droite) : F.E.S = 37 à 82 Hz, F.E.E = 127 Hz, C.G.S = 147 Hz.

Lorsque le spectre possède un pic de niveau, l'allure caractéristique de l'aire entre les D.P.S.C montre une décroissance puis une zone minimale (figure 3.5.4).

Les valeurs de F.E.S dépendent plus de la fréquence du pic que de la largeur de bande du spectre.

Si dans l'exemple précédent, la largeur de bande maximale est ramenée de 500 à 300 Hz avec la même valeur de fréquence pour le pic de niveau alors F.E.S = 36 à 79 Hz, F.E.E = 102 Hz et C.G.S = 106 Hz.

Par nature, F.E.E et C.G.S décroissent alors que la F.E.S est maintenue constante ce qui permet d'avancer que la fréquence d'équilibre spectral peut rendre compte de la forme de l'enveloppe spectrale tous paramètres égaux par ailleurs. La position fréquentielle du pic de niveau modifie les valeurs numériques des trois grandeurs (figure 3.5.5).

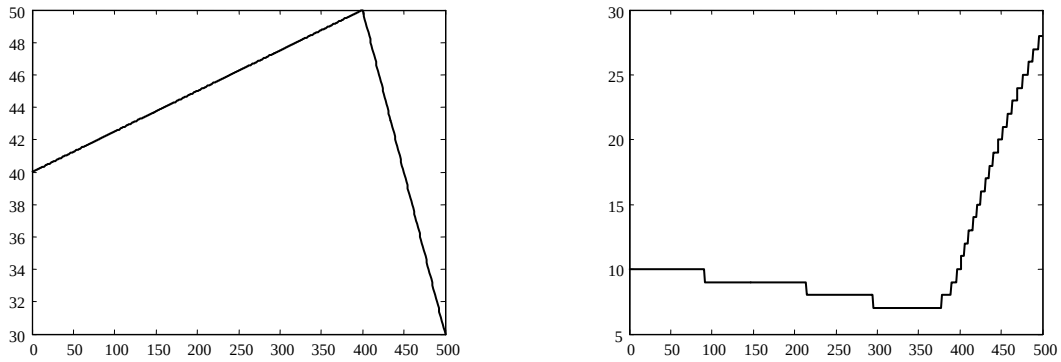


Figure 3.5.5 : Spectre (dB-Hz) avec un pic de niveau (à gauche) et évolution de l'aire entre les D.P.S.C pour chaque fréquence (à droite) : F.E.S = 295 à 377 Hz, F.E.E = 314 Hz, C.G.S = 289 Hz.

Les différences de niveau entre le pic et les bornes fréquentielles du spectre modifient les valeurs des aires et donc celles de la F.E.S. Mais si les écarts de niveau respectifs sont constants, tous les autres paramètres étant égaux, aucune des grandeurs F.E.S, F.E.E et C.G.S n'est modifiée.

Spectre avec trois pics de niveau :

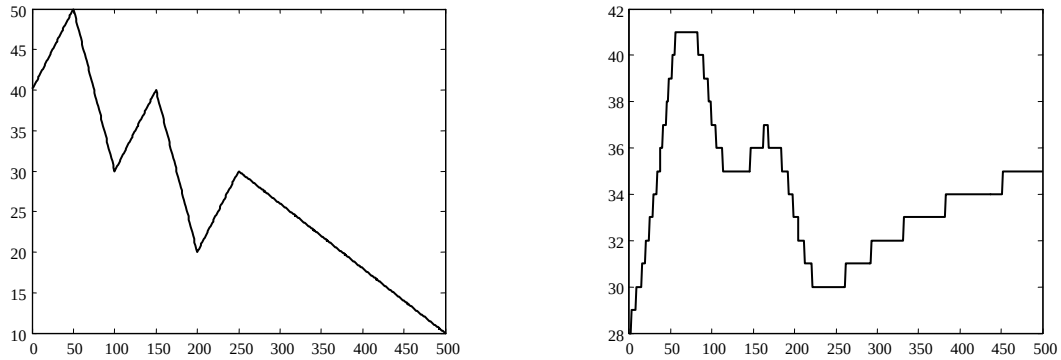


Figure 3.5.6 : Spectre (dB-Hz) avec trois pics de niveau (à gauche) et évolution de l'aire entre les D.P.S.C pour chaque fréquence (à droite) : F.E.S = 1 et 2 Hz, F.E.E = 47 Hz, C.G.S = 57 Hz.

Par analogie avec les enveloppes spectrales de voyelle les niveaux des pics décroissent avec la fréquence. Dans ce cas, comme l'illustre la figure 3.5.6, le minimum de l'aire est obtenu pour les fréquences les plus basses. Son augmentation ne se manifeste que si la fréquence du premier pic est accrue (comportement analogue à celui de la figure 3.5.5). La

figure 3.5.7 illustre le phénomène. On note que les valeurs de F.E.S demeurent dans les fréquences très basses alors que celles de F.E.E et C.G.S sont sensiblement accrues.

Ici également, la mesure de la fréquence d'équilibre spectral semble s'attacher à traduire la forme du spectre. Dans les deux cas précédents, les F.E.S sont constantes et « s'étirent » avec le décalage de la forme spectrale vers les hautes fréquences.

Si la forme de l'enveloppe est rompue, par exemple avec un niveau relatif très élevé du troisième pic par rapport aux deux précédents, alors la F.E.S quitte le domaine des très basses fréquences. C'est l'exemple de la figure 3.5.8. Les valeurs de F.E.S y sont très proches de celles de F.E.E et de C.G.S mais ce n'est le fait que de l'exemple considéré et l'observation n'est pas généralisable.

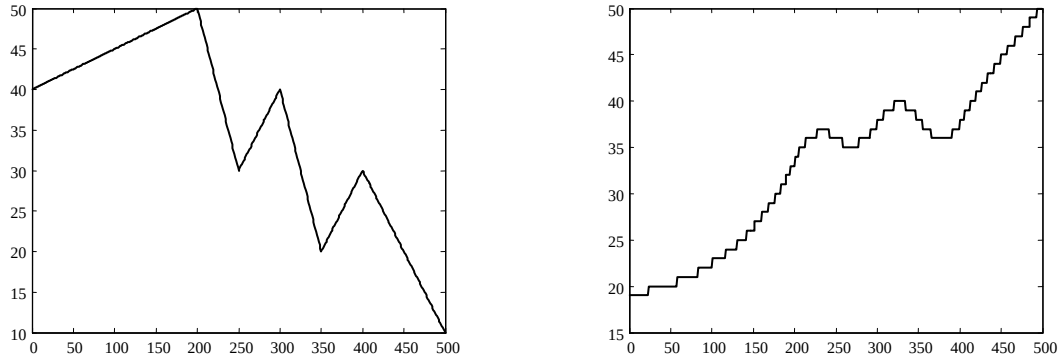


Figure 3.5.7 : Spectre (dB-Hz) avec trois pics de niveau (à gauche) et évolution de l'aire entre les D.P.S.C pour chaque fréquence (à droite) : F.E.S = 1 à 23 Hz, F.E.E = 160 Hz, C.G.S = 151 Hz.

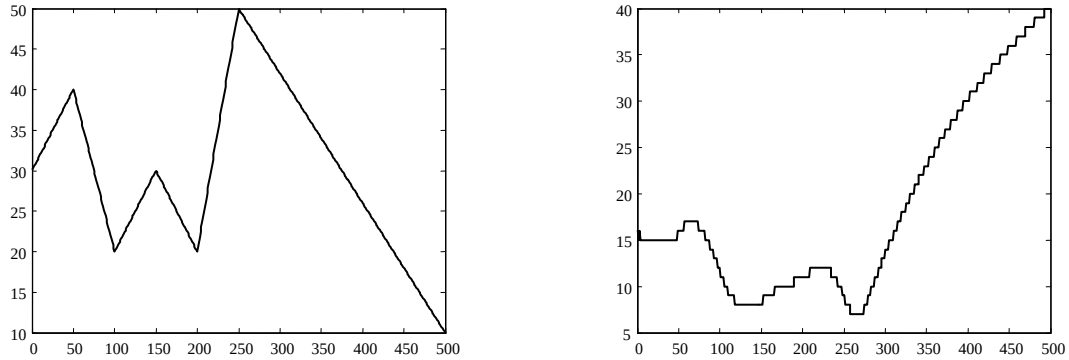


Figure 3.5.8 : Spectre (dB-Hz) avec trois pics de niveau (à gauche) et évolution de l'aire entre les D.P.S.C pour chaque fréquence (à droite) : F.E.S = 258 à 265 Hz, F.E.E = 259 Hz, C.G.S = 251 Hz.

La comparaison des trois paramètres montre que la mesure de F.E.S se distingue bien des deux autres et rend compte d'une approche différente.

F.E.E et C.G.S expriment la mesure d'une fréquence « centrale » dans le spectre alors que F.E.S est à même de traduire la forme de l'enveloppe.

L'application aux voyelles donnera des valeurs très faibles pour les spectres classiques à pente négative avec des niveaux de formants décroissant avec la fréquence. L'existence

d'une F.E.S plus élevée indiquera une rupture de forme spectrale ayant des niveaux relatifs élevés pour les formants d'ordre supérieur (ou pour des bandes hautes fréquence) par rapport au premier formant. Corrélativement, dans les bandes basses fréquence, le niveau du premier formant sera fortement diminué.

La mesure de F.E.S peut donc être un indicateur préalable à des mesures d'énergie ou à la mesure du niveau relatif des formants.

Le résultat est de type dichotomique : valeurs très basse fréquence ou non. F.E.S se présenterait donc comme une mesure seuil évitant de faire un suivi de rapports énergétiques ou des rapports de niveau des formants.

Dans un traitement automatique de multiples spectres, l'apparition de valeurs de F.E.S intermédiaires entre les bornes fréquentielles maximales pourrait indiquer l'existence de formes spectrales inhabituelles (§ 3.3.1 page 74) sans avoir recours à un examen visuel.

Les mesures de F.E.E et de C.G.S ne le permettent pas. Par contre, leur estimation permet à l'expérimentateur de disposer d'indicateurs de suivi énergétique.

Application aux Expérimentations

Préalablement à la définition de la fréquence d'équilibre spectral, au cours des travaux de thèse, un critère a été établi pour étudier la proximité des D.P.S.C basse et haute fréquence.

Il s'agissait de qualifier un spectre d'équilibré lorsque l'écart entre les deux D.P.S.C était en moyenne inférieur à 3dB, signe de deux distributions cumulées proches l'une de l'autre. La séparation en deux bandes était effectuée entre la fréquence du second et du troisième formant (seconde anti résonance). Ce critère a permis de distinguer des spectres prononcés en situation émotionnelle forte dans le cas de l'analyse de l'enregistreur phonique d'un avion accidenté ([3], D.E.A P.Vidal 1994). Par la suite, la F.E.S a été définie et expérimentée sur ce même enregistrement et sur le test de Stroop [4].

Les résultats ont montré que plus de 60% des voyelles prononcées pendant la phase 3 du test de Stroop possèdent des valeurs de F.E.S en dehors du voisinage des bornes du spectre. Pour certaines voyelles, comme le [u] ou le [ø], ce pourcentage moyen est de 77%. Pour l'enregistreur phonique, les chiffres sont moins importants mais les voyelles concernées peuvent être mise en lien avec le contexte émotionnel [4]. Une analyse préliminaire a indiqué que l'évolution de la valeur moyenne de F.E.S par groupes de voyelles supposés a priori exempts de stress avec ceux constitués de phonèmes prononcés dans un contexte émotionnel peut être un indicateur de l'état émotionnel du locuteur (D.E.A P.Vidal, 1994)

Pour toutes les autres expérimentations où la F.E.S a été mise en place et mesurée, de tels résultats ne sont pas apparus. Cet état de fait a été également observé pour la fréquence d'équilibre énergétique et le centre de gravité spectral (exception faite pour ce dernier de l'expérimentation de conduite automobile).

3.5.3. Le « Digital Amplitude Length » (D.A.L)

Sa dénomination résulte de la publication en langue anglaise du paramètre [5].

Le « Digital Amplitude Length » est une estimation de la longueur euclidienne d'un signal numérique observé sur une durée finie. Sa mesure n'est pas propre au signal vocal. On peut rapprocher ce paramètre de la mesure de « plateaux » et « d'épines » dans le signal temporel (Liberman 2003, §5.2.4 page 141).

Dans le cas des signaux de voyelle, leur structure particulière revêt un intérêt pour démontrer l'usage de cette estimation de longueur.

Sur un domaine de fréquence borné, le signal de voyelle comprend un nombre fini d'harmoniques de la fréquence fondamentale conférant au signal temporel une forme particulière et caractéristique. Le modèle théorique lui attribue une allure périodique.

Les pics principaux d'amplitude correspondent aux maxima de pression acoustique atteints après l'ouverture des cordes vocales et chaque période se compose d'une succession de pics d'amplitude moindre. C'est l'allure caractéristique du signal temporel d'un son composé.

En pratique, il se rajoute à cette représentation académique une composante de bruit généralement large bande provenant du souffle d'air et le cas échéant du bruit ambiant présent à la prise de son. Sauf cas particulier, son niveau sonore est faible.

De plus, les signaux réels ne possèdent pas une périodicité stricte : la durée des périodes successives varie. On parle plutôt de signal quasi - périodique. Enfin, l'allure interne à chaque quasi – période n'est pas rigoureusement la même.

Mais un modèle théorique simplifié du signal peut être envisagé et confronté à la mesure du D.A.L pour en étudier les variations.

Signal Sinusoïdal

Soit une fonction sinusoïdale y = sin x. La longueur d'arc L d'une fonction sur (0-π) s'écrit : $L = \int_0^{\pi} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$. Appliquée à la fonction sinus, cette expression devient $L = 2\sqrt{2}.E(\sqrt{1/2})$ où E désigne une intégrale elliptique de seconde espèce telle que $E(k,\phi) = \int_0^{\phi} (1 - k^2.\sin^2\theta)^2\, d\theta$ avec $\Phi = \pi/2$.

La longueur du sinus sur une période vaut donc : $L = 4\sqrt{2}.E(\sqrt{1/2})$.

Soit une fonction sinusoïdale y = A.sin(x/B). L'expression théorique de la longueur euclidienne L d'une période s'écrit alors :

$$L = 4\sqrt{A^2 + B^2}\, E\left(\frac{A}{\sqrt{A^2 + B^2}}\right)$$

En posant B = (1/2.π.f), la longueur de la période d'un son pur de fréquence f et d'amplitude A peut être calculée en opérant la résolution de l'intégrale elliptique par des méthodes de calcul numérique. On trouve L = 4.A, et L = 4.A.P pour P périodes.

Cette longueur ne dépend pas de la fréquence mais seulement de l'amplitude maximale du sinus.

Il s'agit de la longueur du signal sinusoïdal analogique.

Avec une représentation numérique du signal, cette longueur peut être estimée par la somme des distances euclidiennes entre les amplitudes des échantillons successifs. Mais cette distance n'a pas de sens physique puisqu'elle possède deux dimensions : une amplitude et une durée. C'est pourquoi, elle est approchée par la somme des différences d'amplitude, en valeur absolue, entre les échantillons successifs. Cette somme est le « Digital Amplitude Length » ou D.A.L :

$$DAL = \sum_{i=1}^{N-1} DAL(i) = \sum_{i=1}^{N-1} \left| A_i - A_{i+1} \right|$$

dans laquelle N désigne le nombre total d'échantillons d'amplitude respective A_i pour toute la durée du signal observé. D.A.L s'exprime en unités d'amplitude.

En réalisant un échantillonnage du signal sinusoïdal avec M échantillons par période espacés de T_s , la valeur théorique de la longueur L est obtenue quelle que soit la fréquence f pourvu que M>4 (M = 4 décrit un signal triangulaire). La fréquence d'échantillonnage $f_s = 1/T_s$ doit être supérieure à 4 fois la fréquence f du signal sinusoïdal.

Pour f_s = 5.f et A = 1, DAL = 3,8042 ; pour f_s = 10.f et A= 1, DAL = 3,8042 et pour f_s = 20.f, DAL = 4. La règle générale théorique est donc vérifiée.

Même si un nombre réduit d'échantillons suffit à approcher le résultat, l'accroissement de la fréquence d'échantillonnage est recommandé pour tendre vers la valeur théorique du signal analogique.

Signal composé

La méthode de calcul ayant été validée pour le signal sinusoïdal, il est désormais possible de l'appliquer aux signaux plus complexes et en particulier à celui d'un signal numérique composé d'une fréquence fondamentale et d'un nombre fini d'harmoniques.

Compte tenu de son allure temporelle plus complexe, le nombre minimal d'échantillons M_{min} nécessaire au calcul du D.A.L est d'autant plus grand que le nombre K de fréquences est élevé.

Les simulations montrent qu'il est nécessaire de choisir M_{min} au moins égal à 4.K pour obtenir des valeurs du D.A.L proches de la valeur limite obtenue lorsque M devient très grand. En termes de fréquence d'échantillonnage, la préconisation revient à choisir f_s au moins égal à 4.K fois la fréquence fondamentale f (K est le nombre de fréquences : 3 par exemple pour un fondamental et deux harmoniques).

Lorsque le signal composé $\sum_{k=1}^{K} a\sin(2\pi kft)$ $(k \in N^*)$ est numérisé et que le calcul du D.A.L est effectué alors : DAL ≈ 4.A.K.N , ∀ f et ∀ M > M_{min} (N est le nombre de périodes).

Le tableau suivant donne les valeurs du D.A.L pour différentes valeurs des paramètres et pour une seule période de signal. Avec N périodes, les valeurs sont à multiplier par N.

	K = 3	K = 4
A = 0.2 $F_s = 4.K.F_0$ $F_s = 100.K.F_0$	 2,4 2,4629	 3,2 3,2494
A = 1 $F_s = 4.K.F_0$ $F_s = 100.K.F_0$	 12 12,3144	 16 16,2468

Les valeurs exactement conforme à la règle formulée apparaissent pour $f_s = 4.K.F_0$ ou pour $M_{min} = 4.K$ car le signal est en dents de scie et les arcs sont des droites.

Pour des fréquences d'échantillonnage supérieures, les arcs augmentent la longueur totale vers une valeur limite très proche des D.A.L obtenus pour les $f_s = 100.K.F_0$.

Dans le cas général où l'amplitude A_k des harmoniques k n'est pas identique la règle générale de calcul du D.A.L est la suivante : DAL $\approx K.\sum_{k=1}^{K} A_k$ où les A_k sont les amplitudes respectives du fondamental A_1 et des (K-1) autres harmoniques.

Par exemple, pour un signal $\sum_{k=1}^{K} A_k \sin(2\pi kft)$ largement échantillonné avec K = 3 (2 harmoniques) et $A_1 = 4$, $A_2 = 3$, $A_3 = 2$, DAL $\approx 26{,}43$ et la valeur théorique est égale à 27.

Signal de voyelle

Les préconisations précédentes ont montré que le signal devait au moins être échantillonné à 4 fois K la fréquence fondamentale F_o (K est le nombre total de fréquences du signal composé y compris la fréquence fondamentale). Dans le cas d'un signal de voyelle possédant une F_o comprise entre 100 et 300 Hz analysé sur une bande 0-5000 Hz, cette fréquence d'échantillonnage minimale vaut alors 20000 Hz et reste compatible avec les caractéristiques classiques de la numérisation.

Les résultats des simulations précédentes montrent que le D.A.L augmente avec le nombre et le niveau des harmoniques du signal composé, à durée égale, pour une période par exemple.

Ce paramètre peut tester si la « complexité » temporelle du signal de voyelle, sa forme, sont modifiées par des perturbations de la prononciation dues par exemple à des facteurs personnels du locuteur (émotion, fatigue …).

Comme le signal est quasi – périodique il est envisagé de procéder au calcul du D.A.L période par période, c'est à dire depuis un premier maximum d'amplitude jusqu'au suivant, puis d'en obtenir la valeur moyenne pour l'ensemble de la prononciation avec la dispersion correspondante (écart-type, coefficient de variation …).

Trois paramètres peuvent alors être définis par période $T_{o,}$ accompagnés du calcul de leur valeur moyenne sur l'ensemble des périodes de la voyelle:

DAL_{To} qui est le D.A.L pour la période T_o ; $\frac{DAL_{To}}{T_0}$; $\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}$;

Les deux derniers paramètres permettent de rapporter la valeur du D.A.L respectivement à la durée de la période et à sa dynamique (A_{max} et A_{min} sont les amplitudes extrêmes du signal sur la durée d'observation). Ils viennent compléter les mesures classiques du jitter de la fréquence fondamentale et du shimmer.

Pour étudier les fluctuations à court terme de ces grandeurs il est proposé d'estimer un jitter et un « factor jitter » ce qui conduit aux formulations suivantes. Les jitters évaluent les fluctuations à court terme des paramètres. Les « factor jitters » normalisent les jitters correspondants.

$$jitter(DAL_{To}) = \frac{1}{M-1}\sum_{i=1}^{M}\left|(DAL_{To})(i) - (DAL_{To})(i+1)\right| \quad \text{(en unités d'amplitude)}$$

$$factorjitter(DAL_{To}) = 100\,\frac{jitter(DAL_{To})}{< DAL_{To} >} \quad \text{(en \%)}$$

$$jitter\left(\frac{DAL_{To}}{T_0}\right) = \frac{1}{M-1}\sum_{i=1}^{M}\left|\left(\frac{DAL_{To}}{T_0}\right)(i) - \left(\frac{DAL_{To}}{T_0}\right)(i+1)\right| \quad \text{(en unités d'amplitude par seconde)}$$

$$factorjitter\left(\frac{DAL_{To}}{T_0}\right) = 100\frac{jitter\left(\frac{DAL_{To}}{T_0}\right)}{mean\left(\frac{DAL_{To}}{T_0}\right)} \quad \text{(en \%)}$$

$$jitter\left(\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}\right) = \frac{1}{M-1}\sum_{i=1}^{M}\left|\left(\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}\right)(i) - \left(\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}\right)(i+1)\right|$$

(sans dimension)

$$factorjitter\left(\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}\right) = 100\frac{jitter\left(\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}\right)}{mean\left(\frac{DAL_{To}}{|A_{\max} - A_{\min}|_{To}}\right)} \quad \text{(en \%)}$$

Le D.A.L et ses paramètres dérivés renseignent sur de fines modifications de la forme temporelle du signal période par période.

Elles peuvent être dues à des variations du niveau des harmoniques. Elles sont difficilement détectables, période après période, dans le plan spectral soit parce que l'enveloppe spectrale est trop lissée, soit parce que l'estimation du niveau individuel des harmoniques est imprécis compte tenu des effets de la pondération du signal et de la dualité temps – fréquence.

En ce sens, le calcul des longueurs euclidiennes est un complément des méthodes d'analyse spectrale notamment celles qui conduisent à l'obtention d'une enveloppe (méthode L.P.C).

De même, les calculs de rapports d'énergie dans le spectre restent beaucoup trop macroscopiques ou deviendraient trop nombreux pour rendre compte de variations de l'amplitude des harmoniques.

Limitations

● Le calcul du D.A.L est sensible à la présence de bruit qui en accroît la valeur puisqu'il rajoute au signal original, et à tous les instants, de nouvelles amplitudes.

Soit le cas d'un signal composé d'une fréquence fondamentale à 100 Hz d'amplitude 1 et de deux harmoniques de même amplitude et de fréquence 200 et 300 Hz. La figure 3.5.3.1 à gauche montre le signal échantillonné à 10000 Hz et la figure de droite le même signal avec un bruit additif aléatoire (sur une période).

Le calcul du D.A.L conduit aux valeurs respectives suivantes : 12,3144 pour le signal non bruité et 321,5531 pour le signal bruité.

Le niveau du bruit accentue fortement le phénomène et permet de démontrer l'accroissement du D.A.L en présence de bruit.

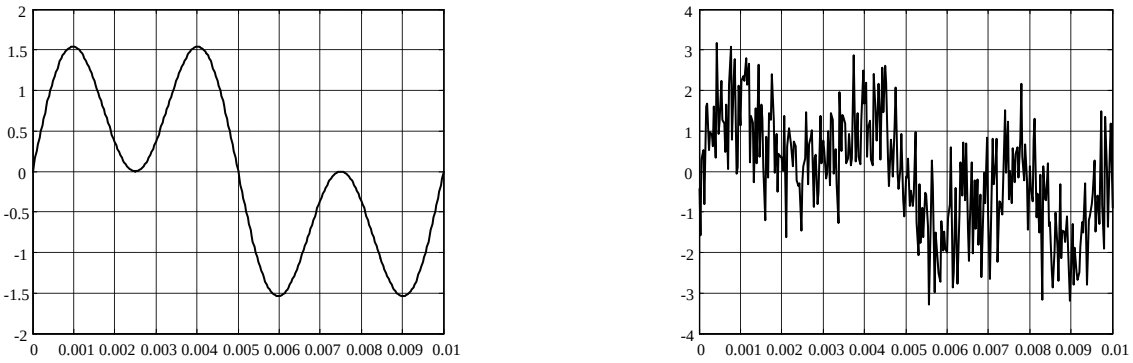


Figure 3.5.3.1 : Signal résultant de la sommation d'une sinusoïde de fréquence 100 Hz avec ses deux premiers harmoniques (de même amplitude = 1) observé sur une période et le même signal avec un bruit additif.

Le D.A.L du signal bruité n'est pas égal à la somme du D.A.L du signal et du D.A.L du bruit. Le calcul des différences successives de niveau pour chaque échantillon s'opère en valeur absolue et conduit à cette inégalité.

Si le bruit provient de la prononciation, son effet n'est pas préjudiciable dans la mesure où il est partie intégrante du signal à analyser et peut être un signe des effets

perturbateurs recherchés. C'est ce que mesure le « Harmonic to Noise Ratio » (H.N.R) dans le cas de voix pathologiques pour lesquelles le niveau relatif de bruit est souvent considéré comme un important paramètre.

Les expressions verbales correspondantes peuvent en effet posséder un niveau de bruit important du par exemple à la fermeture incomplète des cordes vocales. Dans ces cas pathologiques, le H.N.R est élevé. Il devient faible lorsque les perturbations se manifestent plutôt par des vibrations irrégulières des cordes vocales. Dans ce cas, les paramètres comme le shimmer ou le jitter varient également et la présence de bruit est faible.

Pour cette recherche, le H.N.R n'a pour l'instant pas été mesuré en prenant pour hypothèse que les prononciations dans les expérimentations n'ont pas de caractère pathologique pouvant induire la présence notable de bruit.

Même si le H.N.R peut rendre compte d'irrégularités dans la vibration des cordes vocales affectant la largeur de bande et la distribution fréquentielle des harmoniques, le choix d'en témoigner par la mesure du jitter, du shimmer et donc du D.A.L a été préféré.

Ces grandeurs ne nécessitent pas de mettre en œuvre un traitement complexe du signal et opèrent dans le même registre que le H.N.R. Ce choix s'inscrit dans l'option générale de la recherche qui vise à manipuler le moins possible le signal original.

Si le bruit provient de l'ambiance à la prise de son alors la comparaison du D.A.L pour différentes voyelles ne pourra être représentative que si ce bruit conserve les mêmes caractéristiques acoustiques. Dans le cas contraire, la comparaison n'est pas possible sauf à envisager une méthode préalable d'élimination.

● La mesure du D.A.L est par essence liée au niveau du signal acquis. Avant toute analyse, on doit s'assurer qu'aucune modification du niveau autre que celle du locuteur n'est intervenue. Comme pour tous les paramètres liés à l'amplitude, le niveau d'enregistrement et la distance bouche – microphone doivent être maintenus constants. Il est recommandé de normaliser l'amplitude de chaque voyelle segmentée par rapport à sa valeur maximale avant de procéder aux calculs du D.A.L. Cette opération rend alors comparables les mesures effectuées pour chaque phonème et permet de réaliser les calculs de valeurs moyennes par groupe de voyelles.

Application aux Expérimentations

Dans les expérimentations où le D.A.L et ses paramètres associés ont été mesurés (vols court-courriers, conduite automobile, inertie du sommeil, enregistreur phonique), seule la mesure du rapport DAL/T_0 a fourni des variations sensibles et seulement pour les deux premières expériences (§ 3.2.1 page 63, et § 3.2.2 page 67). Elles n'ont pu être observées que lorsque la fréquence fondamentale présentait elle-même une augmentation ou une diminution importante. Ce rapport accroitrait la variation conduisant à isoler davantage le groupe de voyelles concernées.

Il exagère les modifications de la seule fréquence fondamentale permettant de mettre en évidence des variations jusque là peu importantes et en lien avec le contexte perturbant.

Il constitue une voie de recherche pour mettre mieux en évidence les faibles modifications du signal, somme toute plus fréquentes et plus utiles à renseigner sur l'état psychophysiologique, que les importants changements observés dans les situations émotionnellement très marquées où l'apport d'une détection se justifie moins.

3.5.4. La Modulation d'Amplitude des Maxima de Pression des Cycles Vibratoires

Les analyses de signaux de voyelle ont révélé qu'il existait des phonèmes pour lesquels une modulation d'amplitude des pics principaux d'amplitude pouvait apparaitre. La figure 3.5.4.1 montre une voyelle où le phénomène se manifeste. Il s'agit d'une prononciation après un réveil brutal (expérimentation sur l'inertie du sommeil). L'observation a pu être également faite pour l'étude de l'enregistreur phonique.

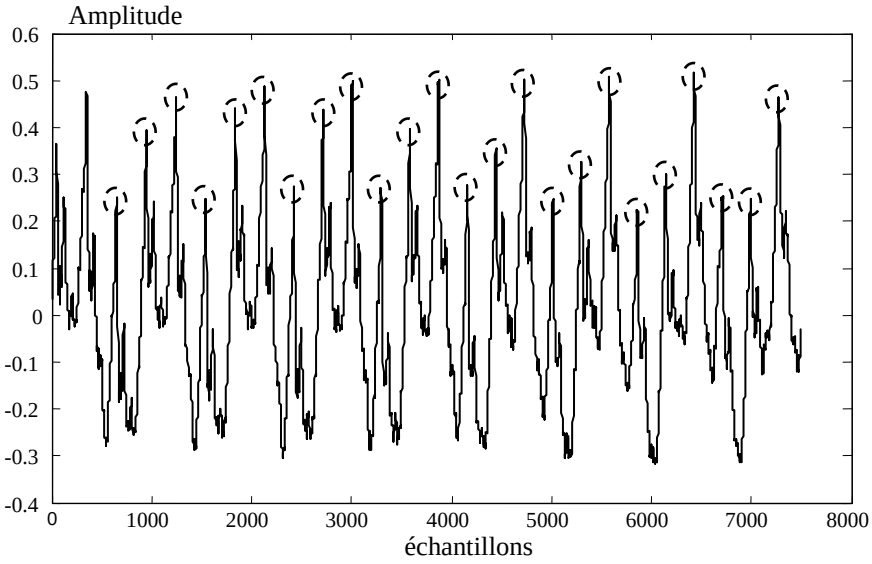


Figure 3.5.4.1 : voyelle [a] extraite d'un mot prononcé consécutivement au réveil du locuteur, F_0 = 153 Hz ; jitter = 1,29 Hz ; shimmer = 3,76 dB. Les cercles indiquent les maxima d'amplitude détectés pour la mesure de la fréquence fondamentale par cycle vibratoire.

Les maxima d'amplitude détectés (repérés par des cercles dans la figure 3.5.4.1) servent au calcul de la période de chaque cycle et donc à la fréquence fondamentale. On en déduit la valeur moyenne sur l'ensemble de la prononciation mais aussi le jitter et le shimmer qui dans ce cas est bien plus élevé qu'en conditions normales de prononciation. Les variations du niveau des pics sont en effet importantes mais elles semblent également présenter une modulation sinusoïdale. Son estimation diffère de la mesure du shimmer. La nature des informations est différente : un shimmer élevé n'est pas nécessairement assorti d'une modulation d'amplitude des pics et l'existence d'une modulation n'est pas nécessairement synonyme d'un shimmer élevé.

Il va donc s'agir de détecter la présence ou non de cette modulation et d'en estimer la fréquence notée f_{mod} .

Cette mesure est à rapprocher de celle d'une fréquence de modulation d'amplitude de la fréquence fondamentale que des micros – tremblements des cordes vocales provoquerait et qui sont à la base d'un appareillage détecteur de mensonge par démodulation d'amplitude (F.H.Fuller 1974, §5.2.4 page 141). Mais contrairement aux résultats des expériences où la modulation de fréquence disparaît avec le stress (P.S.E, C.V.S.A), ici la réponse est plus irrégulière.

Les maxima d'amplitude repérés par des cercles sur la figure 3.5.4.1 sont disposés chronologiquement autour de leur valeur moyenne. Une régression sinusoïdale est calculée sur ces données par les fonctions statistiques de Matlab® (figure 3.5.4.2).

La figure 3.5.4.2 montre une courbe en dents de scie. Elle relie le point représentatif de la différence entre la valeur du maxima d'amplitude et la valeur moyenne de ces maxima au point suivant. La régression sinusoïdale qui présente le plus faible écart avec la courbe en dents de scie s'y superpose.

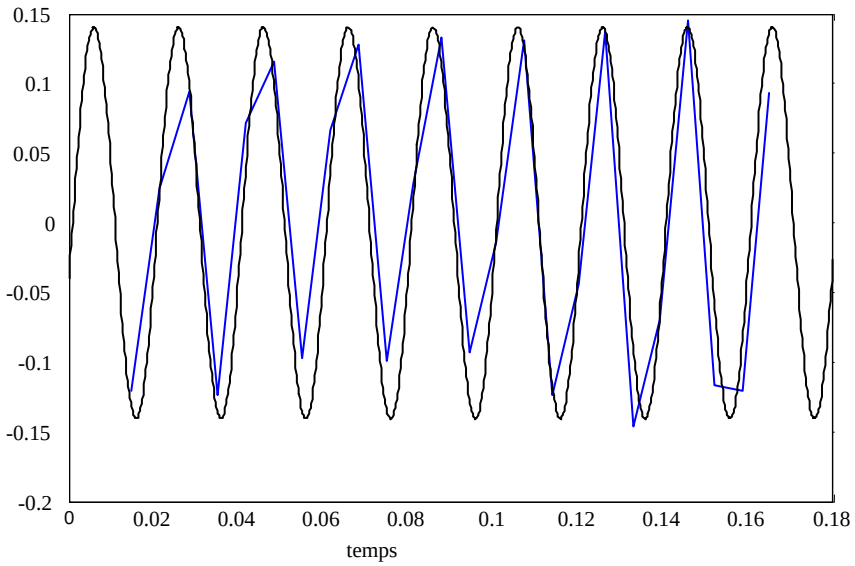


Figure 3.5.4.2 : différence entre la valeur de chaque pic principal de niveau (cercles de la figure 3.5.4.1) et leur valeur moyenne et régression sinusoïdale sur ces données (temps en s).

La fonction sinusoïdale de la régression possède une fréquence de 50,1 Hz sans relation aucune avec la fréquence du courant alternatif du secteur. La qualité de la régression est estimée grâce au coefficient de détermination $R^2 = 0,93$ mais cette qualité est aussi fortement dépendante de la durée du phonème.

En effet, lorsqu'il ne s'agit pas d'étudier des voyelles prononcées isolément mais segmentées dans le discours naturel, même s'il est contrôlé en imposant au locuteur le vocabulaire à prononcer, leur durée peut être très courte. Dès lors, en l'absence de suffisamment de cycles vibratoires, la mesure d'une modulation d'amplitude n'est pas adaptée. Sur l'exemple de la figure 3.5.4.1, 23 cycles vibratoires ont été retenus ce qui confère à la mesure un bon degré de confiance.

Il est difficile de fixer un seuil minimal pour ce nombre. Les mesures fournissent le nombre de cycles, la fréquence de la fonction sinusoïdale de régression et le coefficient de détermination.

Application aux Expérimentations

La mesure de la fréquence de modulation a été confrontée à quatre situations expérimentales : inertie du sommeil, vols court-courriers, conduite automobile et enregistreur phonique.

Pour les trois dernières aucune tendance chronologique n'a pu être observée comme l'illustre la figure 3.5.4.3 pour le commandant de bord de l'avion ayant effectué les rotations quotidiennes court-courrier. Elle représente la valeur numérique de la fréquence de modulation pour chaque voyelle analysée. Sur 92 phonèmes seuls 22 ont un coefficient de détermination inférieur à 0,8 avec des nombres de cycles très variables.

Les premières voyelles sont prononcées en début de journée et les dernières en fin. Il n'apparaît pas de tendance générale à la croissance ou à la décroissance.

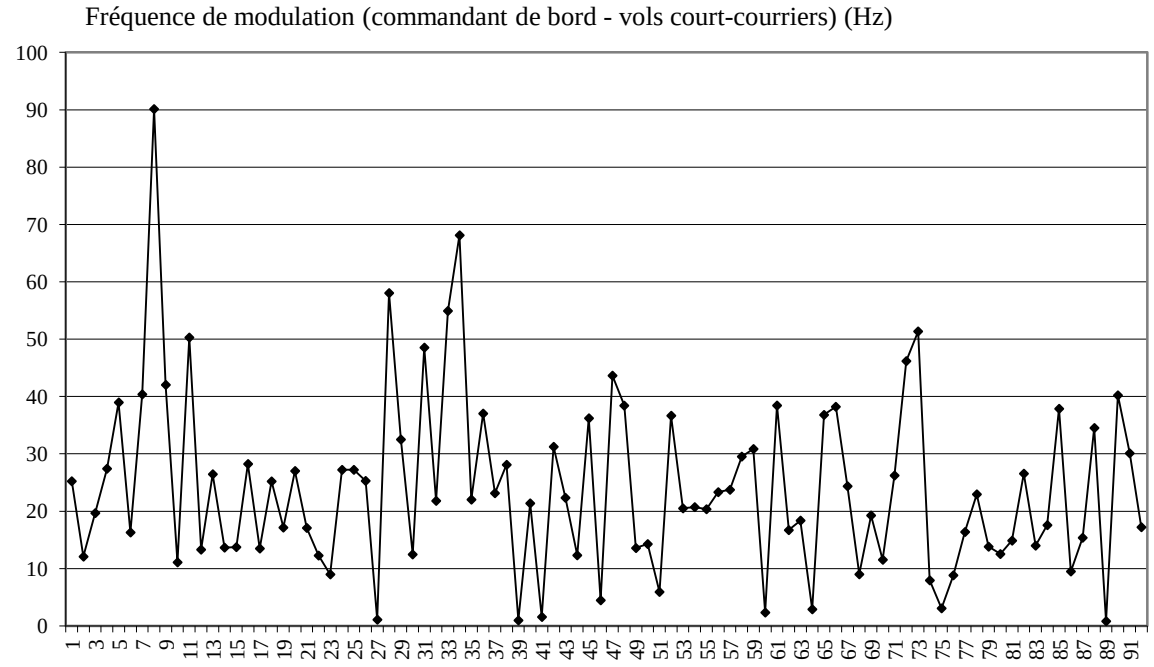


Figure 3.5.4.3 : Fréquence de modulation (Hz) par voyelle (commandant de bord de l'expérimentation sur la fatigue en vols court-courriers).

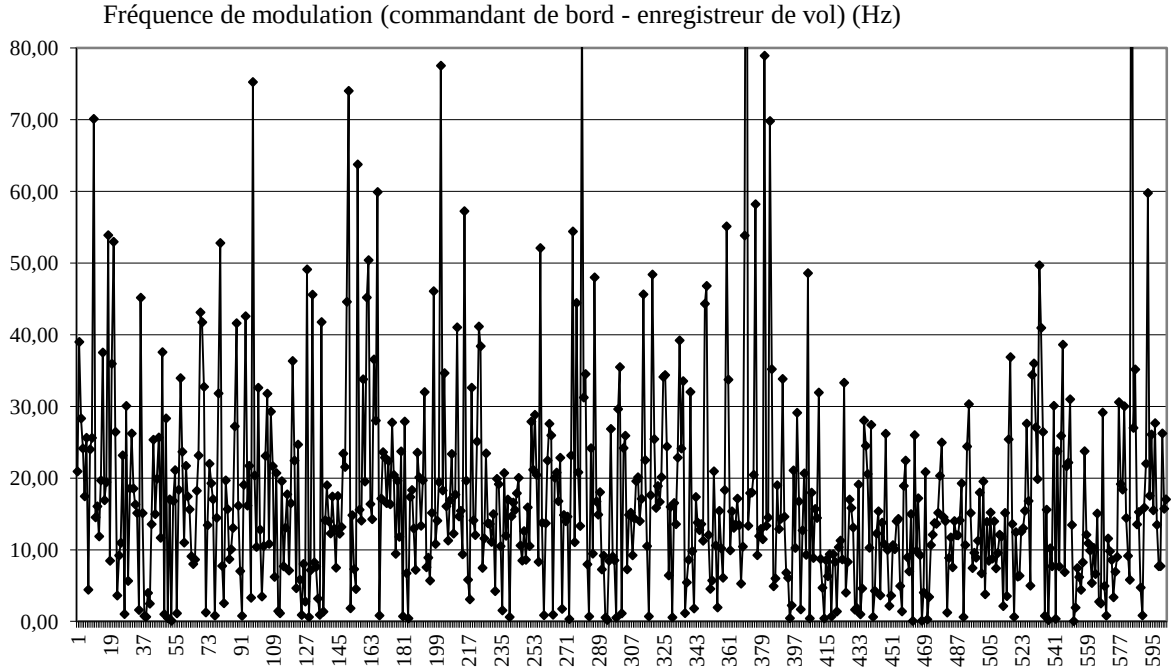


Figure 3.5.4.4 : Fréquence de modulation (Hz) par voyelle (commandant de bord de l'enregistreur phonique).

Pour l'analyse de l'enregistreur phonique, une brusque chute de la fréquence de modulation est repérée en fin de vol pour le commandant de bord (figure 3.5.4.4). La baisse s'étend et couvre presque toute la phase de discussion entre les membres d'équipage après la seconde apparition du problème et avant les dernières opérations avant le décollage.

Pour l'expérimentation sur l'inertie du sommeil la conclusion est différente [24]. L'examen de la figure 3.5.4.5 montre que la fréquence de modulation est plutôt constante. Les 67 premières voyelles (groupe 1) sont issues de prononciations en situation normale et les 40 dernières (groupe 2) sont prononcées après que le locuteur endormi ait été subitement réveillé.

Le coefficient de détermination reste plutôt faible avant la 67ème voyelle et très proche de 1 pour les suivantes (figure 3.5.4.6).

En se donnant un seuil pour R^2 égal à 0.8, très peu de voyelles du groupe 1 émergent. En se donnant une valeur de seuil plus basse, la qualité de la régression diminue d'autant, et ne permet pas de valider l'existence d'une telle fréquence de modulation.

Elle existe après le réveil du locuteur (groupe 2) et n'est semble-t-il pas présente en situation normale de prononciation (groupe 1).

Un autre élément d'analyse intervient pour l'interprétation des résultats. Le nombre de cycles vibratoires est variable (de 5 à 25 environ). Les voyelles du groupe 2 ne possèdent pas un nombre plus élevé de cycles que celles du groupe 1.

Avec un nombre de périodes plus élevé, non seulement R^2 pourrait augmenter mais la fréquence de modulation serait elle aussi susceptible de varier.

La représentation de la fréquence de modulation, ou de R^2, en fonction du nombre de cycle ne montre pas de corrélation apparente entre les grandeurs.

De même, aucune relation n'existe entre la fréquence de modulation et R^2.

L'apparition d'une fréquence de modulation sous l'effet de l'inertie du sommeil semble donc acceptable pour le locuteur testé.

Les deux résultats majeurs obtenus sont une diminution brusque et constante de la fréquence de modulation en présence d'un stress émotionnel (enregistreur phonique) et la validation de son apparition après le brusque réveil d'un locuteur soumis à des exercices psychomoteurs (inertie du sommeil).

Les deux expériences ne sont pas comparables et leurs résultats ne le sont donc pas. Rien ne permet de statuer définitivement en l'absence d'autres investigations et d'autres locuteurs, d'autant que toutes les expériences n'ont pas conduit à de telles observations. Néanmoins, ce paramètre et sa mesure statistique associée présentent par deux fois des propriétés qui peuvent le rendre apte à détecter des troubles du locuteur.

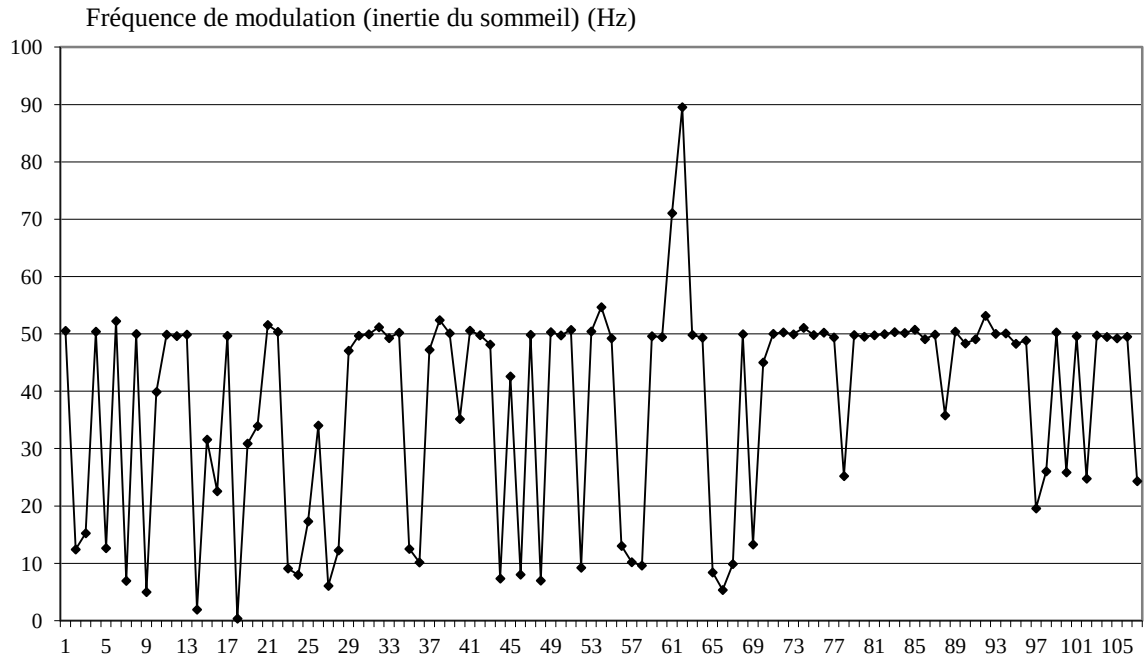


Figure 3.5.4.5 : Fréquence de modulation (Hz) par voyelle (pilote de l'expérimentation sur l'inertie du sommeil) [24].

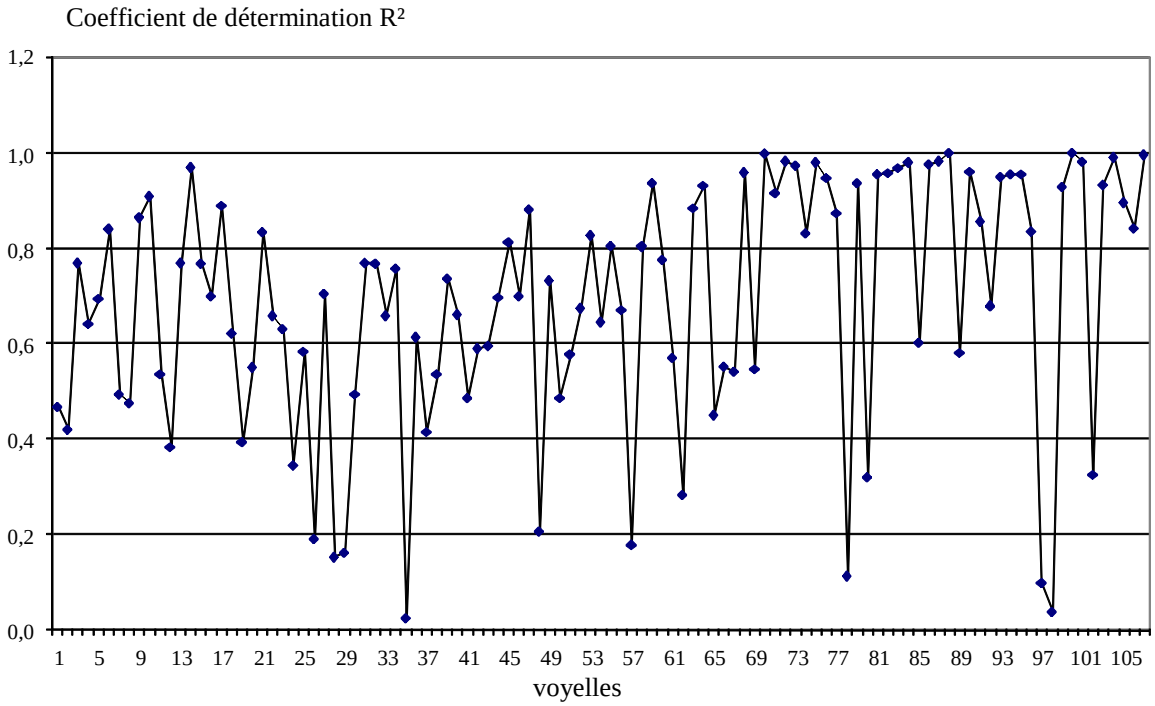


Figure 3.5.4.6 : Coefficient de détermination R² par voyelle (pilote de l'expérimentation sur l'inertie du sommeil) [24].

3.5.5. La Modélisation

La problématique de la modélisation s'appuie sur les résultats du Laboratoire d'Anthropologie Appliquée de l'Université Paris 5 qui indique que la fatigue évolue linéairement avec la durée de service des pilotes (Anthropologie 1998, §5.2.5 page 142). Elle a donc pour objet de rechercher un modèle acoustique traduisant cette évolution chronologique linéaire.

On pose $P_i(\Delta t_j)$ la valeur numérique moyenne du paramètre acoustique P_i pour l'enregistrement j réalisé Δt_j heures plus tard que l'origine des temps. Celle-ci est choisie à l'heure de début du premier enregistrement que l'on note $\Delta t_0 = 0$ h.

L'unité temporelle est exprimée en heures car elle est adaptée à une échelle d'estimation de la fatigue mais elle peut être quelconque.

On emploie le terme enregistrement plutôt que voyelle segmentée, car la valeur du paramètre acoustique peut être une moyenne réalisée sur un ensemble de voyelles prononcées dans des groupes lexicaux voisins. C'est cet ensemble, éventuellement mais pas nécessairement réduit à l'unité, que désigne le terme enregistrement.

On intègre donc l'hypothèse que les manifestations acoustiques des perturbations, si elles existent, ne sont pas instantanées mais apparaissent probablement sur une durée comprenant plusieurs voyelles successives. C'est aussi une hypothèse adaptée à l'étude de la fatigue. Rien n'empêche cependant de construire la modélisation sur des successions de mesures de voyelles à des instants Δt_j précis plutôt qu'à des moyennes sur des intervalles de temps plus grands. Dans ce cas, la modélisation est plutôt adaptée à des perturbations dont les effets se manifestent sur des périodes plus courtes.

Le choix de travailler en temps relatif semble mieux adapté à la recherche d'une évolution chronologique et rend les modèles comparables d'une expérimentation à l'autre.

La problématique de la modélisation revient à trouver les coefficients α_i qui satisfont au système linéaire d'équations suivant :

$$\begin{aligned}
&\alpha_1 P_1(\Delta t_0) + \alpha_2 P_2(\Delta t_0) + \ldots\ldots\ldots + \alpha_n P_n(\Delta t_0) = p \times \Delta t_0 + b \\
&\alpha_1 P_1(\Delta t_1) + \alpha_2 P_2(\Delta t_1) + \ldots\ldots\ldots + \alpha_n P_n(\Delta t_1) = p \times \Delta t_1 + b \\
&\vdots \\
&\vdots \\
&\alpha_1 P_1(\Delta t_m) + \alpha_2 P_2(\Delta t_m) + \ldots\ldots\ldots + \alpha_n P_m(\Delta t_m) = p \times \Delta t_m + b
\end{aligned}$$

dans lequel p désigne la pente du modèle linéaire recherché, b l'ordonnée à l'origine, n le nombre de paramètres acoustiques et m+1 le nombre d'enregistrements (Δt_0 est le premier enregistrement).

Pour éliminer une inconnue dans le problème, l'ordonnée à l'origine b est fixée. Elle ne peut pas être nulle, car alors la première équation du système serait inexacte. Les paramètres acoustiques ne sont pas tous nuls quelle que soit la référence choisie. L'ordonnée à l'origine sera égale à la moyenne des paramètres acoustiques à $\Delta t_0 = 0$ heure (en général le premier enregistrement mais pas nécessairement).

Le nombre d'équations est égal à m+1, c'est-à-dire m enregistrements plus celui à l'origine des temps Δt_0 . Le nombre d'inconnues est égal à n+1. Ce sont les n coefficients α_i plus la pente p.

Sous forme matricielle le système se présente comme suit :

$$\begin{bmatrix} P_1(\Delta t_0) & P_2(\Delta t_0) & \cdots & \cdots & P_n(\Delta t_0) & -\Delta t_0 \\ P_1(\Delta t_1) & P_2(\Delta t_1) & \cdots & \cdots & P_n(\Delta t_1) & -\Delta t_1 \\ \vdots & \vdots & & & \vdots & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ P_1(\Delta t_m) & P_2(\Delta t_m) & \cdots & \cdots & P_n(\Delta t_m) & -\Delta t_m \end{bmatrix} \times \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \vdots \\ p \end{bmatrix} = \begin{bmatrix} b \\ b \\ \vdots \\ \vdots \\ b \end{bmatrix}$$

$$\text{avec } b = \frac{\sum_{i=1}^{n} P_i(\Delta t_0)}{n} \text{ et } \Delta t_0 = 0\,\text{h}.$$

Sous réserve que le déterminant de la matrice des P_i soit non nul, la matrice des inconnues du problème est obtenue par inversion matricielle.

Quand le nombre d'équations est égal au nombre d'inconnues (m = n), la résolution du système aboutit à une pente p qui définit une droite passant exactement par la combinaison linéaire des paramètres acoustiques aux différents temps, et dont l'ordonnée à l'origine vaut b.

C'est l'exemple de la figure 3.5.5.1 avec les deux paramètres $F_0/100$ et jitter pour lesquels on obtient :

- l'équation de la droite (pente et ordonnée à l'origine), c'est-à-dire l'équation du modèle ;
- les deux coefficients α_1 et α_2 respectivement pour Fo/100 et le jitter ;
- la somme des écarts quadratiques qui s'écrit ici :

$$\sum_{\Delta t_0}^{\Delta t_2} ((\alpha_1 P_1(\Delta t_i) + \alpha_2 P_2(\Delta t_i)) - (p\Delta t_i + b))^2$$

Elle est nulle dans le cas où le nombre d'équations est égal au nombre d'inconnues ;

- le carré du coefficient de corrélation R (ou coefficient de détermination). Il vaut 1 dans le cas où le nombre d'équations est égal au nombre d'inconnues ; Il est calculé comme suit pour l'exemple considéré avec deux mesures temporelles en dehors de Δt_0 :

$$R = \frac{\sum_{\Delta t_0}^{\Delta t_2} ((p\Delta t_i + b) \times (\alpha_1 P_1(\Delta t_i) + \alpha_2 P_2(\Delta t_i)))}{\sqrt{\sum_{\Delta t_0}^{\Delta t_2} (p\Delta t_i + b)^2} \times \sqrt{\sum_{\Delta t_0}^{\Delta t_2} (\alpha_1 P_1(\Delta t_i) + \alpha_2 P_2(\Delta t_i))^2}}$$

- le pourcentage de variation du modèle entre le premier et le dernier temps c'est-à-dire :

$$100 \times \frac{(p\Delta t_2 + b) - (p\Delta t_0 + b)}{(p\Delta t_0 + b)}$$

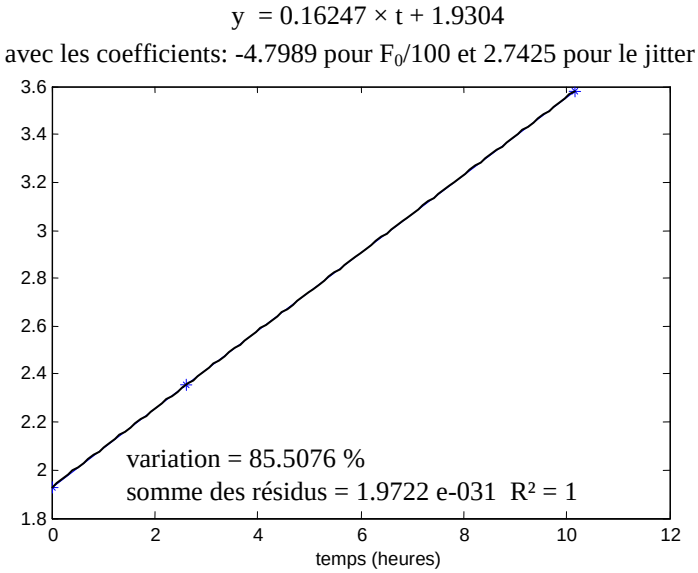


Figure 3.5.5.1 : Exemple de modélisation linéaire avec deux paramètres acoustiques et trois enregistrements.

Quand le nombre d'équations est supérieur au nombre d'inconnues (m > n), les points expérimentaux affectés de leur coefficient respectif ne se distribuent pas exactement sur une droite.

C'est l'exemple de la figure 3.5.5.2 en prenant les deux mêmes paramètres acoustiques (Fo/100 et jitter) mais pour sept équations correspondant à sept enregistrements.

La droîte reste très voisine des points compte-tenu de la valeur de la somme des écarts quadratiques et de la valeur du coefficient de détermination.

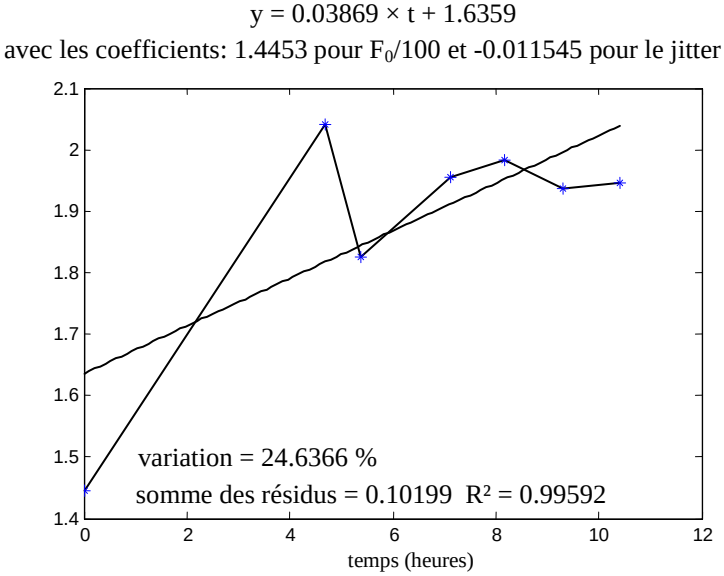


Figure 3.5.5.2 : Exemple de modélisation linéaire avec deux paramètres acoustiques et sept enregistrements.

Quand le nombre d'équations est inférieur au nombre d'inconnues (m<n), le problème n'est pas soluble. Il importe de préciser dans quelles conditions cette situation se présente.

Avec m+1 enregistrements, et pour chacun d'entre eux n paramètres acoustiques, alors si m est inférieur à n, la matrice des inconnues possède une dimension n+1 plus grande que celle du nombre d'équations (à savoir le nombre m+1 d'enregistrements).

Pratiquement, cela signifie qu'avec trois mesures temporelles (m = 2) à Δt_0 , Δt_1 et Δt_2 il existe trois équations et on ne peut rechercher des modèles qu'avec les données mesurées de deux paramètres acoustiques.

Construire une modélisation fondée sur un grand nombre de paramètres acoustiques va impliquer de disposer d'un aussi grand nombre d'enregistrements (+ 1). C'est une considération pratique à prendre en compte pour le dépouillement et la mise en œuvre des expérimentations.

Valeurs conditionnées ou valeurs non conditionnées **:** La figure 3.5.5.3 montre les résultats de la modélisation pour les deux paramètres Fo et jitter, sept enregistrements mais avec un des deux paramètres dont les valeurs sont non conditionnées. Elle est à comparer avec la figure précédente 3.5.5.2 réalisée pour les mêmes paramètres et les mêmes enregistrements mais où les valeurs moyennes par enregistrement de Fo sont conditionnées, c'est-à-dire divisées par 100.

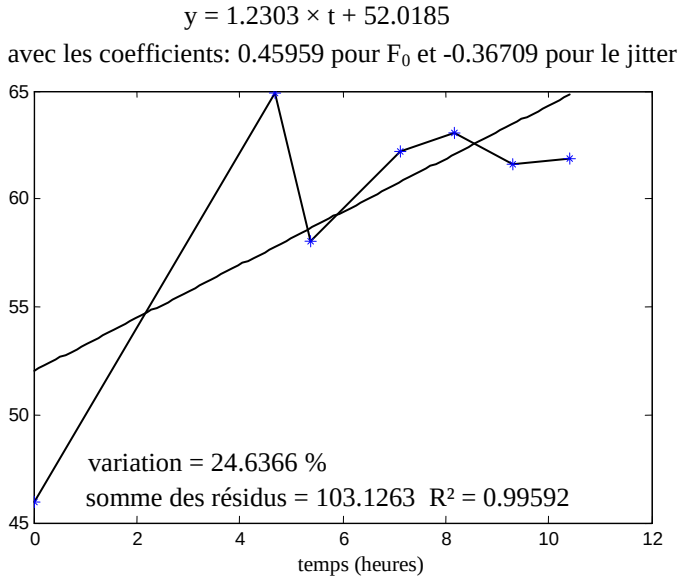


Figure 3.5.5.3 : Exemple de modélisation avec des valeurs non conditionnées.

L'allure des deux figures est exactement la même ainsi que le pourcentage de variation entre la première et la dernière valeur du modèle. L'équation du modèle et les coefficients de la combinaison linéaire des paramètres acoustiques changent puisque la construction du modèle n'utilise pas les mêmes valeurs numériques des mesures. Pour la même raison, il n'existe pas de facteur 100 entre les α_1 des valeurs non conditionnées et celles des valeurs conditionnées.

Peu importe le choix qui est fait. La conséquence du conditionnement ou non des valeurs aboutit à des échelles différentes : entre 1 et 2 dans l'exemple 3.5.5.2 (valeurs conditionnées) et entre 40 et 70 sur la figure 3.5.5.3 (valeurs non conditionnées). Le choix est à la discrétion de l'expérimentateur pour adapter les échelles de variation de chacun des paramètres acoustiques.

Application aux expérimentations lors de vols court-courriers (deux locuteurs) et lors d'une journée de conduite automobile (un locuteur)

Les deux expérimentations s'attachent à étudier la fatigue dans des conditions de transport. Pour l'expérience de vol court-courrier on ne dispose que de trois enregistrements à des heures différentes alors que sept ont été réalisés lors de la journée de conduite automobile.

S'agissant de rechercher un modèle commun aux deux expérimentations parce qu'elles sont toutes deux censées induire une fatigue comparable, il ne pourra être construit qu'avec les mesures de deux paramètres acoustiques. Cette condition est dictée par les considérations pratiques précédentes liées au nombre d'inconnues et au nombre d'équations.

D'où l'obtention des modèles suivants. Ce sont ceux pour lesquels les diverses combinaisons de deux paramètres acoustiques qui ont été testées aboutissent à des équations et des coefficients relativement proches, que ce soit pour deux ou trois locuteurs.

Les modèles à deux paramètres acoustiques qui sont les plus proches pour deux des trois locuteurs :

F_o/100 et shimmer :

conducteur automobile : $y = 0{,}012759.t + 1{,}1904$ avec $\alpha_1 = 0{,}73839$ et $\alpha_2 = 0{,}26483$
pilote OPL : $y = 0{,}011678.t + 0{,}96318$ avec $\alpha_1 = 0{,}63732$ et $\alpha_2 = -0{,}30884$

F_o/100 et $(DAL/T_o)/100$:

conducteur automobile : $y = 0{,}033143.t + 3{,}2928$ avec $\alpha_1 = 4{,}7562$ et $\alpha_2 = -0{,}3011$
pilote OPL : $y = 0{,}037483.t + 3{,}7254$ avec $\alpha_1 = 5{,}5042$ et $\alpha_2 = -0{,}38631$

jitter et $(DAL/T_o)/100$:

conducteur automobile : $y = 0{,}1502.t + 3{,}9109$ avec $\alpha_1 = -0{,}10031$ et $\alpha_2 = 0{,}55436$
pilote OPL : $y = 0{,}14783.t + 4{,}2178$ avec $\alpha_1 = 1{,}5117$ et $\alpha_2 = -0{,}16339$

Les modèles à deux paramètres acoustiques qui sont les plus proches pour les trois locuteurs :

shimmer et $(DAL/T_o)/100$:

conducteur automobile : $y = 0{,}026386.t + 3{,}4654$ avec $\alpha_1 = 1{,}4256$ et $\alpha_2 = 0{,}22072$
pilote CDB : $y = 0{,}094927.t + 3{,}9598$ avec $\alpha_1 = 0{,}67926$ et $\alpha_2 = 0{,}45236$
pilote OPL : $y = 0{,}049221.t + 3{,}5675$ avec $\alpha_1 = 2{,}0717$ et $\alpha_2 = 0{,}30004$

Le nombre important de décimales n'est mentionné que pour comparer les équations et les coefficients. Les pentes et les ordonnées à l'origine peuvent être écrites avec une précision égale au millième.

Ces trois derniers modèles sont ceux qui possèdent les équations dont la pente et l'ordonnée à l'origine sont les plus proches pour les trois locuteurs. C'est la seule combinaison de paramètres acoustiques qui fournisse ce type de résultat.

Un modèle commun aux trois locuteurs peut être proposé pour lequel la pente, l'ordonnée à l'origine et les coefficients α_i sont leur valeur moyenne pour les trois équations précédentes. L'équation est la suivante :

$$y = 0{,}057.t + 3{,}664 \text{ avec } \alpha_1 = 1{,}392 \text{ et } \alpha_2 = 0{,}324$$

La modélisation ne s'applique ici qu'à la fatigue, mais la méthode est transposable à d'autres types de perturbations sous réserve de disposer des tendances psychophysiologiques d'évolution chronologique.

Sa validité dépendra du nombre de locuteurs impliqués et du nombre de mesures réalisées. Avec seulement trois locuteurs et des conditions de fatigue hypothétiquement supposées similaires, les équations proposées ne sont pas généralisables. C'est avant tout le principe de la modélisation qui est introduit pour de futurs travaux.

3.6. Synthèse des Travaux Théoriques

Tous les paramètres acoustiques de la revue de la littérature n'ont pas été testés dans les situations expérimentales mises en œuvre. Néanmoins, le périmètre du signal qu'ils exploitent est repris par les mesures. Par exemple, le champ du non linéaire n'a pas été exploré par l'opérateur de Teager mais par le plus grand exposant de Lyapunov même si les deux mesures ne sont pas comparables. C'est aussi le cas pour les paramètres de la prosodie que les analyses reprennent en partie.

Le compte rendu synthétique des travaux ne fait pas état des essais pour le choix des conditions d'analyse et des méthodes de mesure.

Il n'est pas fait mention des mesures infructueuses comme celles liées aux micros tremblements, aux coefficients auto-régressifs de l'analyse spectrale ou encore aux corrélations entre les paramètres.

La méthode choisie pour mener ces études a été de développer des outils informatiques de laboratoire plutôt que de confier les analyses à des logiciels ou des appareillages spécifiques. La chaîne d'acquisition et de traitement du signal est donc totalement paramétrable et contrôlable. Elle peut réaliser le traitement automatique de grandes listes de fichiers audio de voyelles segmentées.

Les signaux de consonnes n'ont été analysés que pour l'expérimentation sur le modèle de la charge de travail aéronautique d'Airbus. Ils ont été abandonnés par la suite au constat de la grande variabilité de leurs propriétés dynamiques et spectrales même pour des prononciations au repos. Ce fut le cas pour le temps d'établissement du voisement. La durée des structures phonétiques CVC a été étudiée mais il s'agit plus d'une caractéristique de débit que d'un paramètre propre au signal de consonne. Elle présente l'intérêt d'entrevoir de futures études sur les transitions entre phonèmes ou sur des structures linguistiques plus étendues que celle de la voyelle.

Les dimensions du signal de voyelle ont été explorées dans tous les plans qu'elles définissent à l'exception de la force sonore.

Le niveau sonore des prononciations n'a pas été examiné en tant que tel. La cause de cette absence tient aux contraintes aéronautiques. Des correcteurs automatiques de gain sont présents dans la chaîne d'acquisition et nuisent à l'estimation du niveau. La dimension amplitude est tout de même présente à travers les mesures de shimmer, du D.A.L ou de plusieurs paramètres spectraux.

Les mesures de débit phonatoire ont été occasionelles hormis pour l'expérience sur le réflexe visio phonatoire (durée des structures CVC) et celle de conduite automobile.

Deux raisons à cela : tout d'abord, l'écoute attentive des enregistrements ne permet pas de détecter des signes de variations sensibles, même dans les situations émotionnelles réelles et extrêmes. Bien que dans une situation réelle on se trouve dans un cas opposé à celui du journaliste décrivant l'embrasement du Hindenburg (cf page 7). Les signes émotionnels sont inaudibles (sauf à la fin du vol).

La deuxième raison émane des spécialistes du dépouillement des enregistreurs de vol. Au cours des activités que nécessitent le pilotage, les échanges verbaux entre pilotes et les communications avec le contrôle aérien semblent posséder une variabilité de débit importante pouvant masquer les modifications dues au facteur émotionnel. Dans ces conditions, l'exploration du débit apparaît difficile.

Comme pour le niveau sonore cela ne signifie pas l'abandon de toute recherche en la matière. Il s'agit d'engager une réflexion pour orienter convenablement les mesures afin de tenir compte de ces observations.

Les variations combinées de plusieurs paramètres peuvent rendre compte de l'apparition ou de l'évolution de perturbations vocales. Des combinaisons ont été examinées pour les deux expériences sur la fatigue parce que les études psychophysiologiques laissent à penser que son évolution chronologique est linéaire à l'échelle de la durée de l'expérimentation. Elles ont conduit à élaborer les premiers modèles de variation acoustique chronologique présentant un caractère linéaire. Le nombre de sujets examinés ne les destine pas à être généralisés pour l'instant.

D'un point de vue théorique, les nouveaux paramètres proposés peuvent venir compléter l'ensemble des caractéristiques existantes. Mais leur application aux situations expérimentales n'a pas aboutit à des conclusions qui permettraient de les associer plus directement aux perturbations du locuteur comme c'est le cas pour la fréquence fondamentale. Il est nécessaire d'y apporter des développements complémentaires ou de les limiter à un rôle potentiellement indicatif, en association avec d'autres.

D'un point de vue auditif, la voix peut être caractérisée par sa force, sa hauteur, son timbre et son débit. Pour la force, la hauteur et le débit il existe une mesure acoustique objective correspondante : niveau sonore, fréquence fondamentale, mots par minute par exemple.

Le timbre ne possède pas cette propriété. La forme spectrale en est une composante. Pour autant, la fréquence fondamentale y est inscrite mais c'est sa dimension temporelle qui la caractérise le mieux et offre la plus grande précision pour son estimation.

Il est donc apparu intéressant de se pencher sur les brèves et fluctuantes variations temporelles du signal vocal entre les points de mesure de la fréquence fondamentale. Elles pourraient posséder l'attribut temporel de la composante spectrale du timbre à l'image de ce que sont les périodes fondamentales pour la hauteur sonore. De ces fluctuations temporelles où prend naissance l'aspect « timbral » du son (en complément de son attaque), la mesure de leur longueur euclidienne est la première étape d'une caractérisation.

Un lien avec la forme spectrale n'est pas à exclure comme lorsqu'on rajoute des harmoniques à un son pur. Cette dernière a été examinée par l'étude des distributions cumulées du niveau sonore. L'étude simultanée des deux formes temporelles et spectrales peut être envisagée.

L'exploration de la portion du signal temporel entre les impulsions principales d'amplitude ouvre une voie de recherche. On peut penser à d'autres mesures de longueur fractionnant temporellement la période principale, ou issues de signaux filtrés en fréquence, ou encore de mesures de taux de variation locaux de l'amplitude du signal. Des mesures de ce type peuvent être combinées à celles de la fréquence fondamentale et de ses paramètres dérivés.

L'estimation de la modulation d'amplitude relève aussi de l'intérêt porté à la forme du signal temporel.

Elle résulte d'erreurs de détection de la fréquence fondamentale. En parcourant le signal entre ses maxima, des valeurs de jitter trop élevées ont été mesurées. Ceci était du à l'existence de variations importantes du niveau sonore d'un cycle vibratoire à l'autre. La fréquence fondamentale moyenne de la voyelle était très peu affectée par les mesures biaisées cycle à cycle. Il a donc été nécessaire de corriger la méthode pour comptabiliser tous les pics nécessaires à une estimation correcte des périodes fondamentales successives.

Une modulation d'amplitude semblait en être la cause. Elle a conduit à l'élaboration du paramètre proposé.

Cette constatation montre le défaut de l'automatisation des traitements des signaux rendue pourtant nécessaire si l'on veut disposer d'une base de données expérimentale suffisante. Cette automatisation aurait pu conduire à examiner avec beaucoup moins de soin les diagrammes amplitude–temps et niveau–fréquence du signal. C'est un des enseignements de ces travaux.

Enfin, même si une grande variabilité peut être attendue, la forme de l'attaque pourrait être étudiée tout comme celle des périodes temporelles de transition entre les phonèmes.

Les travaux théoriques sur le signal acoustique se sont attachés à proposer des paramètres plus globalement en lien avec la forme d'onde et la forme spectrale plutôt que d'introduire des caractéristiques qui n'en rendent pas compte comme la fréquence fondamentale ou les formants.

Une méthode de modélisation des variations chronologiques de la combinaison de plusieurs paramètres est introduite.

Ce sont autant d'éléments à développer et à approfondir.

4. CONCLUSION

Le manuscrit a présenté une **synthèse des travaux** pour la détection acoustique des perturbations psychophysiologiques puis développé les éléments de la contribution appliquée au secteur aéronautique. Cette conclusion va s'attacher à en discuter la teneur, à formuler des éléments critiques et à décrire succintement les travaux en cours. Mais auparavant, l'essentiel des apports de la revue générale des connaissances et de ceux des études engagées va être résumé.

Les points à retenir de l'analyse bibliographique sont les suivants.

Le signal vocal est incontestablement modifié lorsque des perturbations psychophysiologiques affectent un locuteur et ce, de manière interculturelle.

Les travaux ont très majoritairement concerné les émotions, en particulier celles que les psychologues qualifient de primaires, qu'elles aient été simulées par des acteurs ou repérées dans des situations réelles.

Les études sur l'analyse, la synthèse et la reconnaissance automatique de ces émotions sont nombreuses et aboutissent à de bons résultats.

Les autres types de causes modificatrices du signal de parole ont été beaucoup moins explorées. C'est le cas pour la fatigue. Le secteur des transports et ses potentielles aptitudes à provoquer émotion et fatigue chez les conducteurs ou les pilotes n'a que peu été traité. Pour la détection de la fatigue, des travaux sont en cours dans le domaine ferroviaire et aérien.

Les études menées sur des émotions bien déterminées ont conduit à définir les familles de paramètres acoustiques sensibles au phénomène. Un grand nombre de paramètres a été testé. Il ressort que les paramètres de la prosodie sont les plus réactifs, mais ils ne sont pas suffisants pour garantir une détection fiable, voire pour définir des degrés dans la réaction individuelle. Les mesures liées au débit de la parole et au spectre en montrent également la sensibilité et en constituent les compléments nécessaires.

D'autres paramètres, plus confidentiels et originaux, ont été introduits dans les multiples travaux mais sans pouvoir conclure sur leur réelle capacité, dans la mesure où ils ont été le fruit d'études ponctuelles spécifiques et sur des échantillons de population réduits.

Ces mesures complémentaires ne permettent pas d'établir des critères ou des seuils de variation. Sauf dans le cas des émotions clairement identifiées par la nomenclature des psychologues, pour lesquelles des profils de vecteurs acoustiques ont pu être proposés, il n'existe pas d'équivalent dès lors qu'elles sont plus subtiles ou que les modifications du signal sont dues à d'autres facteurs.

Par ailleurs, la reconnaissance automatique, dont il faut rappeler qu'elle est fondée sur une analyse du signal, opte pour des procédés qui traitent un large ensemble de mesures, sans

que parallèlement, les travaux sur l'analyse acoustique aient abouti à bien préciser la nature et le degré des variations observées.

L'identification des modifications est réalisée même s'il reste à découvrir de nouvelles caractéristiques plus à même de rendre compte du phénomène. Les études sur l'analyse de la parole ne paraissent pas être parvenues à leur terme dans la mesure où aucun modèle n'existe. Il n'est pas possible de prévoir le type et l'importance de la perturbation à partir des mesures acoustiques comme il est coutume de le rechercher dans les divers domaines de la Physique.

La contribution à ces recherches corrobore l'existence de manifestations vocales singulières pour les situations expérimentales mises en œuvre, qu'elles soient réelles ou simulées.

Le secteur aéronautique a été le principal domaine d'application de l'étude avec l'analyse de l'enregistreur phonique d'un appareil accidenté, des simulations de vol et des enregistrements en vol réel court- et long- courrier. Le domaine de la psychologie a été approché par deux expérimentations, l'une source de stress cognitif et l'autre visant à mettre en évidence un réflexe visio-phonatoire en lien avec l'émotion. Ils ont donné lieu au co-encadrement d'étudiants de Master 1 et 2 et de Doctorat dont les résultats des travaux sont signalés dans le manuscrit et mêlés aux résultats personnels.

Au plan aéronautique, les expérimentations de simulation complètent un petit nombre de travaux similaires recensés dans la littérature. L'analyse détaillée de l'enregistreur phonique revêt quant à elle un caractère plus original tout comme celles des vols réels court- et long- courriers. Les résultats de toutes ces expériences contribuent à une meilleure connaissance de l'impact vocal des conditions d'exercice du métier de pilote (inertie du sommeil, vigilance, fatigue, stress) et des dysfonctionnements qui peuvent apparaître à bord d'un appareil.

La singularité des ces expériences a conduit à procéder en premier lieu à l'examen du comportement des caractéristiques acoustiques classiques du signal vocal comme la fréquence fondamentale, ses paramètres de variation locale et le spectre avec ses grandeurs propres. Comme l'étude bibliographique le mentionne pour divers types de perturbation, la sensibilité de bon nombre de ces grandeurs a pu être confirmée et établie pour ces situations aéroanutiques. Les conclusions des travaux ne consistent pas à comparer les résultats des situations expérimentales étant entendu qu'elles ne touchent pas aux mêmes types de perturbations : stress émotionnel, stress cognitif, fatigue, vigilance, inertie du sommeil (K.R.Scherer, 2002).

Hormis la féquence fondamentale, sensible dans quasiment tous les cas, les autres grandeurs ne varient pas systématiquement de manière notable et significative. On retrouve là une des conclusions de la revue des connaissances. Par exemple, il a été observé des modifications des distances spectrales en relation avec la fatigue matinale lors de vols court-courriers. Pour cette même expérience, un accroissement constant du plus grand exposant de Lyapunov apparaît au cours de la journée de vol d'un des pilotes, et une moyenne plus élevée existe pour les deux navigants à la première escale matinale où des signes de fatigue ont été

observés. La voix d'un conducteur automobile au cours d'une longue journée de conduite a montré une augmentation sensible du centre de gravité spectral. La hausse importante du jitter et la diminution de l'exposant de Lyapunov apparaissent en relation avec l'inertie du sommeil d'un pilote. En revanche, aucune variation sensible des formants n'a été mise en évidence.

Trois paramètres originaux ont été proposés: la fréquence d'équilibre spectral, le « Digital Amplitude Length » et une mesure de modulation d'amplitude. Ils tentent de répondre quantitativement à des questionnements issus de l'observation de spectres et de signaux de voyelles aux caractères inhabituels.

La fréquence d'équilibre spectral est la fréquence pour laquelle la distribution cumulée du niveau sonore dans les parties basse et haute fréquence du spectre qu'elle partage sont les plus proches. Son élaboration provient de spectres de voyelles où des modifications de la répartition énergétique semblent se produire lorsque des effets perturbateurs se manifestent. Elle se distingue d'une fréquence d'équilibre énergétique et d'un centre de gravité spectral par le fait qu'elle peut rendre compte de la rupture de la forme spectrale usuelle des voyelles. Sans dispenser l'expérimentateur de la mesure des premières, elle peut apporter un complément d'information.

Le « Digital Amplitude Length », dont le nom résulte de sa publication en langue anglaise, est une estimation de la longueur, au sens euclidien, que parcourt le signal au cours d'une période fondamentale. D'autres paramètres en découlent pour exprimer ses variations locales. La théorie du chaos en a été l'inspiratrice. L'hypothèse est que la perturbation peut accroître le désordre du signal ou le simplifier. Dès lors, le nombre de vallées et de pics peut varier en quantité et en longueur de signal ce dont le paramètre peut rendre compte. Il peut compléter les mesures de l'exposant maximal de Lyapunov.

En observant les voyelles dans le plan dynamique, une modulation d'amplitude est apparue sur les maxima d'amplitude du signal lorsque le locuteur était perturbé par certaines conditions expérimentales. Elle a provoqué des erreurs de mesure de la fréquence fondamentale. Celle-ci utilise la détection de ces maxima. Ils peuvent posséder un niveau plus faible que d'autres pics du signal suite à la présence de la modulation et entrainer un biais dans l'estimation de la durée des périodes fondamentales. Ces erreurs n'ont pas affecté le calcul de la fréquence fondamentale moyenne car elles n'ont eu lieu que sur quelques périodes mais le jitter l'a traduit par un accroissement important de ses valeurs. Dès lors, l'estimation des périodes fondamentales a été corrigée et une mesure de cette modulation a été effectuée.

On retrouve là l'expression de préoccupations similaires relevées dans la littérature à propos des micros-tremblements des cordes vocales pour la détection du mensonge. Mais ici la méthode est plus directe: le signal ne subit aucun traitement, il reste à l'état brut, puisque la mesure va consister à appliquer une régression non linéaire sinusoïdale dans le plan dynamique.

Les résultats de la confrontation de ces trois nouveaux paramètres avec les situations expérimentales mises en œuvre n'a pas permis d'observer des variations systématiques et significatives en présence de toutes les perturbations. Des variations singulières sont apparues, un peu comme pour bon nombre d'autres paramètres acoustiques. Le résultat est plutôt positif et encourageant afin que ces caractéristiques puissent en tout ou partie acquérir le statut de paramètre du signal vocal apte à être sensible aux phénomènes étudiés.

Les deux dernières témoignent en outre qu'il est possible de prendre appui sur le signal non traité pour extraire des grandeurs sensibles. L'intérêt est aussi de réaliser des mesures peu couteuses en temps et en complexité de calcul.

La contribution comporte aussi une modélisation de l'évolution temporelle de la fatigue par des paramètres acoustiques. L'analyse critique des travaux antérieurs soulève ce manque. Il importait de poser les bases de cette voie de recherche.

À l'appui de travaux sur le sommeil et la vigilance, l'évolution de la fatigue peut être considérée comme linéaire sur les périodes temporelles des vols court-courriers. Il a été entrepris de trouver la combinaison de paramètres acoustiques optimisant une évolution linéaire entre le début et la fin de la journée de travail.

La méthode de calcul aboutit à des équations de droîtes combinant certains paramètres acoustiques. Elles ne sont pas définitives car les données expérimentales inter-locuteur sont en nombre insuffisant. En revanche, une méthode est proposée pour être testée à plus grande échelle et un pas est effectué vers la modélisation des évolutions.

D'un point de vue critique, des seuils de variation ou des critères n'ont pas été proposés et le nombre de paramètres examinés est moins important que tous ceux de la littérature. Plus positivement, de nouvelles situations expérimentales ont été mises en place et le choix des paramètres acoustiques mesurés, qu'ils soient connus ou originaux, couvre une large étendue dans les plans dynamique, spectral et des phases ; enfin, les premiers éléments d'une modélisation sont introduits.

Toutes les procédures de calcul de ces grandeurs sont intégrées dans des programmes informatiques de laboratoire qui autorisent le contrôle du paramétrage des méthodes de mesure et permettent de disposer d'outils parfaitement connus. Cette stratégie présente l'avantage de rendre la recherche indépendante de considérations logicielles dédiées, offre une maîtrise plus complète de tous les éléments de la chaîne d'acquisition et de traitement mais présente l'inconvénient de travailler avec des outils encore peu éprouvés.

D'autres points critiques sont à formuler.

En premier lieu, il faut mentionner le faible nombre de locuteurs testés inhérent à la complexité de construire une expérimentation de plus grande envergure dans le secteur aéronautique. En conséquence, la recherche s'est plutôt orientée vers l'identification de paramètres sensibles que vers la proposition de valeurs seuils plus directement dépendantes de la taille de l'échantillon pour être représentatives.

Le panorama des indicateurs potentiels n'est pas exhaustif. Il a fallu opérer des choix dans la vaste gamme que propose la littérature. Ils ont été guidés par le désir de couvrir l'ensemble des dimensions sonores en ne retenant que les caractéristiques qui semblaient les plus pertinentes. Il manquerait certainement à compléter la liste par la mesure d'au moins un

paramètre du débit phonatoire, de l'opérateur de Teager et du rapport entre le niveau des harmoniques et celui du bruit.

Les seules mesures de niveau sonore qui apparaissent dans l'étude sont celles du shimmer. Des mesures de niveau moyen ont pourtant été réalisées pour contrôler l'absence de corrélation entre les hausses de la fréquence fondamentale moyenne et celles du niveau sonore. Mais comme dans les situations aéronautiques réelles un correcteur automatique de gain opère sur la chaîne d'enregistrement des conversations, l'exploitation du niveau n'est pas pour l'instant pertinente.

Les travaux entrepris revêtent donc un caractère exploratoire et peu d'équipes travaillent sur le sujet dans le secteur aéronautique.

Les perspectives qui se dessinent pour le développement des travaux sont autant d'éléments critiques à l'encontre des travaux réalisés.

Une segmentation automatique des voyelles serait nécessaire pour renforcer l'automatisation existante de la mesure des grandeurs acoustiques sur de grands ensembles de signaux. Elle est réalisée manuellement et l'absence d'automatisation, si elle garantit la conformité des segments prélevés, rend difficile l'adaptation des analyses à de grands corpus de phonèmes. Il est envisagé de construire de nouvelles expérimentations à plus grande échelle, en ciblant précisemment un type de perturbation, pour à la fois rendre réalisable la recherche de seuils ou de critères mais aussi pour développer la modélisation. Ceci impliquera peut-être de corréler les mesures acoustiques à des mesures physiologiques (électroencéphalographie par exemple) ou psychologiques.

Une amélioration de l'interface homme – machine des programmes de calcul faciliterait également le dépouillement des enregistrements et les mesures les plus classiques devront être confrontées à celles de logiciels dédiés à l'analyse du signal vocal.

En complément de quelques paramètres à introduire dans les mesures, la perspective d'une analyse sémantique du discours par des spécialistes de la question pourrait permettre de cibler les périodes temporelles à analyser prioritairement. Des tests subjectifs d'écoute pourraient également y contribuer. Ce sont autant d'appuis possibles pour réaliser la modélisation, en particulier pour l'exploitation des cas réels.

Enfin, il manquerait à préciser les liens entre les variations des grandeurs acoustiques et le fonctionnement de l'appareil phonatoire lorsque leur identification sera achevée et surtout, que des choix définitifs de paramètres auront été retenus. On peut penser aux relations entre les modifications spectrales et des effets visco-thermiques (salivation, température locale), entre les fluctuations à court-terme de la fréquence fondamentale et des phénomènes vibratoires (tension des muscles, élasticité des cordes vocales) et à des considérations aéroacoustiques (respiration, turbulences).

Les applications aéronautiques se poursuivent tant il est vrai que le secteur des transports exprime de plus en plus le besoin d'estimer l'état émotionnel ou de fatigue des conducteurs de véhicules transportant des passagers. Ces estimations ont pour but de prévenir

les erreurs humaines. Ils participent à la compréhension de l'origine des accidents. L'intérêt est manifeste en matière de sécurité aérienne puisqu'aujourd'hui les facteurs humains sont responsables de plus de 75% des accidents. Par ailleurs, leur identification, voire l'estimation de leur degré d'influence, pourront fournir d'utiles informations sur les circonstances lorsqu'elles seront mises en regard des paramètres de vol.

Dans les années à venir, les recherches relatives à la détection des perturbations psychophysiologiques devraient connaitre un essor important. L'analyse de la voix peut y prendre part car elle est non invasive, et que les métiers concernés utilisent nécessairement et naturellement la parole dans leur activité et surtout, parce que la preuve est faite que la voix est un support de leur manifestation.

5. RÉFÉRENCES BIBLIOGRAPHIQUES

5.1. Publications de l'Auteur : Il s'agit uniquement des publications qui relèvent de la thématique du manuscrit.

Revues Internationales

[1] R.Ruiz, C.Legros, A.Guell, Voice analysis: application to the study of the influence of a workload, *J. Acoustique*, 3 (2), pp. 153-159, 1990.

[2] R.Ruiz, C.Legros, A.Guell, Voice analysis to predict the psychological or physical state of a speaker: a future reality, *Aviation Space and Environ. Med.*, 61, pp. 266-271, 1990.

[3] R.Ruiz, C.Legros, The cumulative spectral probability diagram: theory and experiments, *Acustica united with Acta Acustica*, 2 (3), pp.215-222, 1994.

[4] R.Ruiz, E.Absil, B.Harmegnies, C.Legros, D.Poch, Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions, *Speech Communication*, 20 (1-2), pp. 111-129, 1996.

[5] R.Ruiz, P.Plantin de Hugues, C.Legros, Advanced voice analysis of pilots to detect fatigue and sleep inertia, *Acustica united with Acta Acustica*, 96, pp. 567-579, 2010.

Congrès avec Actes

[6] R.Ruiz, C.Legros, Psycho-physiological state influence on acoustic voice characteristics, Acts of the 28th Acoustic Conf. on Physiological Acoustics, Psychoacoustics, Acoustics of Music and Speech, Strbske Pleso, Tchécoslovaquie, pp. 147-150, 1989.

[7] R.Ruiz, C.Legros, Analyse de la voix de pilotes en situation de stress dans un simulateur de vol, Colloque de Physique, colloque C2, vol. 51 (2), pp. 527-529, 1er Congrès Français d'Acoustique, 1990.

[8] R.Ruiz, C.Legros, Estimation of emotional disorders by cockpit voice recorder analysis, American Speech Language Hearing Association Annual Convention, Atlanta, 1 page, U.S.A, 1991.

[9] B.Gramatica, R.Ruiz, C.Legros, Modifications de la fréquence fondamentale de la voix de pilotes: incidents réels et simulés, Journal de Physique IV, colloque C1, vol. 2, pp. 335-338, 2nd Congrès Français d'Acoustique, 1992.

[10] C.Legros, R.Ruiz, Estimation of workload by voice analysis of pilots, 14th International Congress on Acoustics, Beijing, Chine, pp G6-5 (2 pages), 1992.

[11] B.Gramatica, R.Ruiz, C.Legros, A parameter for the evaluation of emotional trouble: the cumulative distribution of amplitude in the speech spectrum, Proc. E.S.C.A Workshop on Speech Processing in Adverse Conditions, Cannes-Mandelieu, France, pp. 235-238, 1992.

[12] R.Ruiz, C.Legros, Estimation de la largeur de bande des formants à partir de la distribution cumulée du niveau sonore dans le spectre, Journal de Physique IV, colloque C5, vol. 4, pp. 461-464, 3ème Congrès Français d'Acoustique, 1994.

[13] B.Gramatica, R.Ruiz, C.Legros, Comparison between time-scale and time frequency approaches for the detection of stress in speech, 15th Int. Congress on Acoustics, Trondheim, Norvège, pp. 49-52, 1995.

[14] R.Ruiz, C.Legros, Vowel spectral characteristics to detect vocal stress, 15th Int. Congress on Acoustics, Trondheim, Norvège, pp. 141-144, 1995.

[15] R.Ruiz, E.Absil, B.Gramatica, B.Harmegnies, C.Legros, D.Poch, Spectrum-related variabilities in stressed speech under laboratory and real conditions, Proc. of the E.S.C.A-N.A.T.O Tutorial and Research Workshop on Speech under Stress, Lisbonne, Portugal, pp. 53-56, 1995.

[16] E.Absil, B.Gramatica, B.Harmegnies, C.Legros, D.Poch, R.Ruiz, Time-related variabilities in stressed speech under laboratory and real conditions, Proc. of the E.S.C.A-N.A.T.O Tutorial and Research Workshop on Speech under Stress, Lisbonne, Portugal, pp. 37-40, 1995.

[17] B.Gramatica, R.Ruiz, Comparaison de deux méthodes de caractérisation du spectre de parole pour l'identification du stress, Actes du 4ème Congrès Français d'Acoustique, vol. 1, pp. 361-364, Marseille, 1997.

[18] S.Vautier, R.Da Silva Neves, C.Martinot, R.Ruiz, Sensibilité de la version française du stai-y à des images avec contenu émotionnel, (résumé), Journal de Thérapie Comportementale et Cognitive, 10, 13, 2000.

[19] R.Ruiz, R.Da Silva Neves, C.Martinot, S.Vautier, Interactions between emotional content of pictures and acoustic features of speech, 17th International Congress on Acoustics, Roma, Italy, 6A 14.04, pp 276-277, 2001.

[20] R.Ruiz, R.Da Silva Neves, C.Martinot, S.Vautier, Vers un réflexe visio-phonatoire, Actes du 6ème Congrès Français d'Acoustique, Lille, pp. 763-766, 2002.

[21] R.Ruiz, C.Legros, P.Plantin de Hugues, Analysing cockpit and laboratory recordings to determine fatigue levels in pilots' voices, Congrès conjoint des sociétés américaine, européenne et française d'acoustique, Acoustics'08, june 29-july 4, pp. 621-625, 2008.

[22] R.Ruiz, P.Plantin de Hugues, C.Legros, Voice analysis as a significant parameter of car driver's fatigue, Proc. of the Acoustics High Tatras 2009, 34th Int. Acoustical Conference – European Acoustical Association (E.A.A), 2009.

[23] R.Ruiz, P.Plantin de Hugues, C.Legros, Analyse de la voix d'un conducteur automobile : effets acoustiques de la fatigue, Actes du 10ème Congrès Français d'Acoustique, CFA 2010, Lyon, 2010.

[24] R.Ruiz, P.Plantin de Hugues, C.Legros, Amplitude modulation of vowel glottal pulses – application to sleep inertia, Actes du 12ème Congrès Français d'Acoustique, CFA 2012, Nantes, 2012.

Rapports de Contrats

[25] C.Legros, R.Ruiz, Étude de l'influence de la charge de travail sur les caractéristiques acoustiques de la voix, 1. Étude bibliographique, Université Toulouse III – Laboratoire d'Acoustique de Métrologie et d'Instrumentation / Centre National d'Études Spatiales, 65 pages, 1988.

[26] C.Legros, R.Ruiz, Étude de l'influence de la charge de travail sur les caractéristiques acoustiques de la voix, 2. Expérimentations en simulateur de vol, Université Toulouse III – Laboratoire d'Acoustique de Métrologie et d'Instrumentation / Centre National d'Études Spatiales, 43 pages, 1990.

[27] B.Gramatica, R.Ruiz, C.Legros, Expérimentations acoustiques à bord des vols minimum crew sur Airbus A340, rapport final, Université Toulouse II – Laboratoire d'Acoustique de l'Université Toulouse-Le Mirail / Services Techniques de la Navigation Aérienne, 65 pages, 1993.

[28] C.Legros, R.Ruiz, Étude de l'influence d'une période de sommeil sur la voix de pilotes d'avion, Rapport final de l'étude B.E.A N°3304/2005, 19 pages, 2007.

[29] R.Ruiz, Prise en compte des spécificités électroacoustiques des communications avion – tour de contrôle pour l'adaptation des mesures de laboratoire sur le signal vocal, phase 2, Rapport final de l'étude B.E.A N°3841/2007, 18 pages, 2007.

[30] R.Ruiz, C.Legros, Étude de l'influence de la fatigue sur les caractéristiques acoustiques de la voix d'un équipage au cours d'une journée de vol court-courrier avec plusieurs rotations, Rapport final de l'étude B.E.A N°3661/2008, 66 pages, 2008.

[31] R.Ruiz, C.Legros, Étude de l'influence d'une période de sommeil sur la voix des pilotes d'avion – paramètres spectraux, Complément au rapport final de l'étude B.E.A N°3304/2005, 27 pages, 2009.

[32] R.Ruiz, C.Legros, Étude de l'influence de la fatigue sur les caractéristiques acoustiques de la voix d'un équipage au cours d'une journée de vol court-courrier avec plusieurs rotations – analyse du cockpit voice recorder, Complément au rapport final de l'étude B.E.A N°3661/2008, 14 pages, 2009.

[33] R.Ruiz, C.Legros, Mesures de la dynamique non linéaire du signal vocal : portrait de phase et exposant de Lyapunov, Complément au rapport final des études B.E.A N°3304/2005 et N°3661/2008, 17 pages, 2009.

[34] R.Ruiz, C.Legros, Étude de l'influence de la fatigue sur les caractéristiques acoustiques de la voix au cours d'une journée de conduite automobile (Phases 1 et 2), Rapport final de l'étude B.E.A N° 3632/2009, 36 pages, 2010.

[35] R.Ruiz, C.Legros, Analyse acoustique des conversations de l'équipage d'un avion à partir de son enregistreur phonique (CVR), Rapport interne du B.E.A, 47 pages, 2010.

5.2. Publications de la Revue Générale des Connaissances

5.2.1. Analyse de la Voix dans le Secteur des Transports

▪ J.Armentrout, D.A.Holland, K.J.O'Toole, W.R.Ercoline, Fatigue and related human factors in the near crash of a large military aircraft, *Aviation Space, and Environ. Med*., 77 (9), pp. 963-970, 2006.

▪ P.Benson, Analysis of the acoustic correlates of stress from an operational aviation emergency, Proc. of the ESCA-NATO tutorial and research workshop on speech under stress, Lisbon, Portugal, September 14-15, pp. 61-64, 1995.

▪ H.Bořil, S.O.Sadjadi, T.Kleinschmidt, J.H.L.Hansen, Analysis and detection of cognitive load and frustration in drivers' speech, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 502-505, 2010.

▪ M.Brenner, M.Cash, Speech analysis as an index of alcohol intoxication – the Exxon Valdez accident, *Aviation, Space and Environ. Med.*, 65, pp. 21-26, 1994.

▪ D.A.Cairns, J.H.L.Hansen, Nonlinear analysis and classification of speech under stressed conditions, *J. Acoust. Soc. Am*., 96 (6), pp. 3392-3400, 1994.

▪ J.J.Congleton, W.A.Jones, S.G.Shiflett, K.P.Mcsweeney, R.D.Huchingson, An evaluation of voice stress analysis techniques in a simulated AWACS environment, *Int. Journal of Speech Tech*., 2, pp. 61-69, 1997.

▪ R.Fernandez, R.W.Picard, Modeling drivers' speech under stress, *Speech Communication*, 40 (1-2), pp 145-159, 2003.

▪ D.M Freund, R.R.Knipling, A.C.Landsburg, R.R.Simmons, G.R.Thomas, A holistic approach to operator alertness research, Proc. of the transportation research board 74th meeting, January 22-28, Washington, D.C, paper n°950636, 1995.

▪ G.R.Griffin, C.E.Williams, The effects of different levels of task complexity on three vocal measures, *Aviation, Space, and Environ. Med.*, 58, pp. 1165-70, 1987.

▪ M.Hagmuller, E.Rank, G.Kubin, Can stress be observed by analyzing the human voice ?, Proc. IEEE INO workshop, 2004.

▪ J.H.L.Hansen, Morphological constrained feature enhancement with adaptative cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect, *IEEE Trans on Speech and Audio Processing*, 2 (4), pp. 598-614, 1994.

▪ J.H.L.Hansen, Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition, *Speech Communication*, 20, pp. 151-173, 1996.

▪ J.H.L.Hansen, B.D.Womack, Feature analysis and neural network-based classification of speech under stress, *IEEE Trans. on Speech and Audio Processing*, 4 (4), pp. 307-313, 1996.

▪ J.H.L.Hansen, C.Swail, A.J.South, R.K.Moore, H.Steeneken, E.J.Cupples, T.Anderson, C.R.A Vloeberghs, I.Trancoso, P.Verlinde, The impact of speech under "stress" on military speech technology, NATO Project4 report AC/232/IST/TG-01, March 2000.

▪ L.Khachatur'yants, L.Grimak, Cosmonaut's emotional stress in space flight, NASA Technical translation TT F-14, 654, 1972.

▪ G.P.Krueger, Y.F.Jones, US Army aviation fatigue-related accidents, 1971-1977, USAARL Report N° 79-1, October 1978.

▪ I.Kuroda, O.Fujiwara, N.Okamura, N.Utsuki, Method for determining pilot stress through analysis of voice communication, *Aviation, Space and Environ. Med*., 47 (5), pp. 528-533, 1976.

▪ V.Kuznetsov, E.Lapayev, Voices in orbit, NASA technical translation TT F-16499, 1975.

▪ F.Loyau, Les émotions dans la communication parlée du pilote de chasse, Actes des VIIèmes RJC Parole, 5-6 Juillet 2007, Paris, France, pp. 100-103, 2007.

▪ M.Nevile, M.B.Walker, a context for error using conversation analysis to represent and analyse recorded voice data, Australian Transport Safety Bureau, Aviation research report B2005/0108, June 2005.

▪ V.B.Maxwell, S.M.Roxborough, Stress in Aviation, *Aviation Space and Environ. Med*., special issue , 56 (7), 1985.

▪ National Transportation Safety Board, Aircraft accident report, NTSB/AAR-99/01, Washington, DC, 1994.

▪ H.J.Older, L.L.Jenney, Psychological stress measurement through voice output analysis, NASA-CR-141723, March 1975.

▪ A.Protopapas, P.Lieberman, Effects of vocal F_0 manipulations on perceived emotional stress, Proc. of the ESCA-NATO tutorial and research workshop on speech under stress, Lisbon, Portugal, September 14-15, pp. 1-4, 1995.

▪ A.Protopapas, P.Lieberman, Fundamental frequency of phonation and perceived emotional stress, *J. Acoust. Soc. Am*., 101 (4), pp. 2267-2277, 1997.

▪ M.A.Rahurkar, J.H.L.Hansen, J.Meyerhoff, G.Saviolakis, M.Koenig, Frequency band analysis for stress detection using a Teager energy operator based feature, 7^{th} Int. Conf. on spoken language processing (ICSLP2002), Denver, USA, September 16-20, vol 3, pp. 2021-2024, 2002.

▪ M.A.Rahurkar, J.H.L.Hansen, Frequency distribution based weighted sub-band approach for classification of emotional/stressful content in speech, Interspeech 2003, Geneva, Switzerland, pp. 721-724, 2003.

▪ I.Saito, O.Fujiwara, N.Utsuki, C.Mizumoto, T.Arimori, Hypoxia-induced fatal aircraft accidents revealed by voice analysis, *Aviation, Space, and Environ. Med*, 51 (4), pp. 402-406, 1980.

▪ S.J.Schneider, M.Alpert, Voice measures of workload in the advanced flight deck: additional studies, NASA Contractor report 4258, November 1989.

▪ B.Schuller, M.Wimmer, D.Arsic, T.Moosmayr, G.Rigoll, Detection of security related affect and behavior in passenger transport, Proc. Interspeech 2008, Brisbane, Australia, pp. 265-268, 2008.

▪ P.Shahidi, S.C.Southward, M.Ahmadian, Real-time estimation of temporal word boundaries without linguistic knowledge, In Proc. IEEE/ASME Joint Rail Conference, JRC2008-63066, April 22-23, Wilmington, Delaware, USA, 2008.

▪ P.Shahidi, S.C.Southward, M.Ahmadian, Fuzzy analysis of speech metrics to estimate conductor alertness, Proc. ASME 2009 rail transportation division fall conference, RTDF2009-18031, October 19-21, Fort Worth, Texas, USA, 2009.

▪ P.Shahidi, S.C.Southward, M.Ahmadian, A holistic approach to estimating crew alertness from continuous speech. In Proc. IEEE/ASME Joint Rail Conference, JRC2010-36211, April 2010.

▪ P.Shahidi, A.R.Soltan, S.C.Southward, M.Ahmadian, Estimating changes in speech metrics indicative of fatigue levels, Proc. ASME 2010 rail transportation division fall technical conference, RTDF2010-42010, October 12-13, Roanoke, Virginia, USA, 2010.

▪ P.Shahidi, S.C.Southward, M.Ahmadian, Estimating crew alertness from speech, in Proc. IEEE/ASME, Joint Rail Conference, JRC 2010-36261, April 2010.

▪ K.Shiomi, S.Hirose, Fatigue and drowsiness predictor for pilots and air traffic controllers, 45th ATCA conf., Oct 22-26, Atlantic city, USA, pp. 1-4, 2000.

▪ P.V.Simonov, M.V.Frolov, Utilization of human voice for estimation of man's emotional stress and state of attention, *Aerospace Med*., 44 (3), pp. 256-258, 1973.

▪ P.V.Simonov, M.V.Frolov, Analysis of the human voice as a method of controlling emotional state: achievements and goals, *Aviation Space, and Environ. Med*., 48 (1), pp. 23-25, 1977.

▪ J.A.Tanford, D.B.Pisoni, K.Johnson, Novel scientific evidence of intoxication: acoustic analysis of voice recordings from Exxon Valdez, *J. of Criminal Law and Criminology*, 82 (3), pp. 579-609, 1992.

▪ J.A.Veltman, A.W.K.Gaillard, Physiological indices of workload in a simulated flight task, *Biological Psychology*, 42, pp. 323-342, 1996.

▪ J.Whitmore, S.Fisher, Speech during sustained operations, *Speech Communication*, 20, pp. 55-70, 1996.

▪ C.E.Williams, K.N.Stevens, On determining the emotional state of pilots during flight: an exploratoty study, *Aerospace Med*., 40 (12), pp. 1369-1372, 1969.

▪ C.E.Williams, K.N.Stevens, Emotions and speech: some acoustical correlates, *J. Acoust. Soc. Am*, 52 (4), pp. 1238-1250, 1972.

▪ C.E.Williams, K.N.Stevens, Vocal correlates of emotional states, *in Speech Evaluation in psychiatry*, edited by J.K.Darby, Grune & Stratton, 1981.

▪ B.D.Womack, J.H.L.Hansen, Classification of speech under stress using target driven features, *Speech Communication*, 20, pp. 131-150, 1996.

▪ B.D.Womack, J.H.L.Hansen, N-Channel hidden markov models for combined spressed speech classification and recognition, *IEEE Trans. on Speech and Audio Processing*, 7 (6), pp. 668-677, 1999.

▪ G.Zhou, J.H.L.Hansen, J.F.Kaiser, Methods for stress classification: nonlinear TEO and linear speech based features, Proc. IEEE Int. Conf. on Acoustics, Speech and signal processing ICASSP'99, 15-19 March, Phoenix, USA, vol 4, pp. 2087-2090, 1999.

▪ G.Zhou, J.H.L.Hansen, J.F.Kaiser, Nonlinear feature based classification of speech under stress, *IEEE Trans. on Speech and Audio processing*, 9 (3), pp. 201-216, 2001.

5.2.2. Voix et Émotion

▪ Â.Abelin, J.Allwood, Cross linguistic interpretation of emotional prosody, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 110-113, 2000.

▪ J.C.Acosta, N.G.Ward, Responding to user emotional state by adding emotional coloring to utterances, Interspeech 2009, 6-10 September, Brighton, UK, pp. 1587-1590, 2009.

▪ H.Ai, D.J.Litman, K.Forbes-Riley, M.Rotaru, J.Tetreault, A.Purandare, Using system and user performance to improve detection in spoken tutoring dialogs, Interspeech 2006, September 17-21, Pittsburgh, USA, pp. 797-800, 2006.

▪ M.Alpert, S.J.Schneider, Voice-stress measure of mental workload, NASA report N°88-23381, 1987.

▪ K.Alter, E.Rank, S.A.Kotz, U.Toepel, M.Besson, A.Schirmer, A.D.Friederici, Accentuation and emotions – two different systems ?, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 138-142, 2000.

▪ K.Alter, E.Rank, S.A.Kotz, U.Toepel, M.Besson, A.Schirmer, A.D.Friederici, Affective ancoding in the speech signal and in event-related brain potentials, *Speech Com*munication, 40, pp. 61-70, 2003.

▪ A.Álvarez, .I.Cearreta, J.M.López, A.Arruti, E.Lazkano, B.Sierra, N.Garay, Application of feature subset selection based on evolutionary algorithms for automatic emotion recognition in speech, ITRW on nonlinear speech processing (NOLISP 07), Paris, France, May 22-25, pp.71-74, 2007.

▪ N.Amir, S.Ron, Towards an automatic classification of emotions in speech, 5th Int. Conf. on Spoken Language Processing (ICSLP 98), Sydney, Australia, November 30 – December 4, paper 0199, 1998.

▪ N.Amir, S.Ron, N.Laor, Analysis of an emotional speech corpus in Hebrew based on objective criteria, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 29-33, 2000.

▪ N.Amir, O.Kerret, D.Karlinski, Classifying emotions in speech: a comparison of methods, Eurospeech 2001, September 3-7, Aalborg, Denmark, pp. 127-130, 2001.

▪ N.Amir, S.Ziv, R.Cohen, Characteristics of authentic anger in Hebrew speech, Eurospeech 2003, Geneva, Switzerland, pp. 713-716, 2003.

▪ C.Araki, S.Hashimukai, S.Motomani, M.Mori, S.Taniguchi, S.Kato, An experimental system to analyze or synthetize baby's emotional voice using the varidation of pitch frequencies, IEEE 2008 Int. Symp. on Universal Communication, pp. 494-498, 2008.

▪ Y.Arimoto, S.Ohno, H.Iida, Acoustic features of anger utterances during natural dialog, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2217-2220, 2007.

▪ Y.Arimoto, H.Kawatsu, S.Ohno, H.Iida, Emotion recognition in spontaneous emotional speech for ananimity-protectd voice chat systems, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 322-325, 2008.

▪ V.Aubergé, L.Lemaître, The prosody of smile, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 122-126, 2000.

▪ V.Aubergé, M.Cathiard, Can we hear the prosody of smile ?, *Speech Communication*, 40, pp. 87-97, 2003.

▪ V.Aubergé, N.Audibert, A.Rilliard, Acoustic morphology of expressive speech: what about contours ?, Speech Prosody 2004, Nara, Japan, March 23-26, pp. 201-204, 2004.

▪ N.Audibert, V.Aubergé, A.Rilliard, The prosodic dimensions of emotion in speech : the relative weights of parameters, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 525-528, 2005.

▪ N.Audibert, D.Vincent, V.Aubergé, A.Rilliard, O.Rosec, Dimensions acoustiques de la parole expressive: poids relatif des paramètres resynthétisés par Pratt vs. LF-ARX, Actes des XXVIes Journées d'Études sur la Parole, Dinard, pp. 341-344, 2006.

▪ N.Audibert, Morphologie prosodique des expressions vocale des affects : quel timing pour le décodage de l'information émotionnelle ?, Actes des VIIèmes RJC Parole, Paris 5-6 Juillet 2007, pp. 12-15, 2007.

▪ N.Audibert, V.aubergé, A.Rilliard, Emotions actées vs spontanées: variabilité des compétences perceptives, 27èmes Journées d'Etudes sur la Parole, 9-13 Juin, Avignon, France, 2008.

▪ C.Baber, B.Mellor, R.Graham, J.M.Noyes, C.Tunley, Workload and the use of automatic speech recognition : the effects of time and resource demands, *Speech Communication*, 20, pp. 37-53, 1996.

▪ R.Banse, K.R.Scherer, Acoustic profiles in vocal emotion expression, *J. of Personality and Social Psychology*, 70 (3), pp. 614-636, 1996.

▪ T.Bänziger, K.R.Scherer, The role of intonation in emotional expressions, *Speech Communication*, 46, pp. 252-267, 2005.

▪ T.Bänziger, D.Grandjean, K.R.Scherer, Emotion recognition from expressions in face, voice, and body: the multimodal emotion recognition test (MERT), *Emotion*, 9 (5), pp. 691-704, 2009.

▪ R.Barra, J.M.Montero, J.Macias-Guarasa, J.Gutiérez-Arriola, J.Ferreiros, J.M.Pardo, On the limitations of voice conversion techniques in emotion identification tasks, Interspeech 2007, August27-31, Antwerp, Belgium, pp. 2233-2236, 2007.

▪ R.Barra-Chicote, F.Fernandez, S.Lufti, J-M.Lucas-Cuesta, J.Macias-Guarasa, J.M.Montero, R.San-Segundo, J.M.Pardo, Acoustic emotion recognition using dynamic Bayesian networks and multi-space distributions, Interspeech 2009, 6-10 September, Brighton UK, pp. 336-339, 2009.

▪ A.Batliner, K.Fisher, R.Huber, J.Spilker, E.Nöth, Desperately seeking emotions or: actors, wizards, and human beings, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 195-200, 2000.

▪ A.Batliner, K.Fischer, R.Huber, J.Spilker, E.Nöth, How to find trouble in communication, *Speech Communication*, 40, pp. 117-143, 2003.

▪ G.Beller, N.Obin, X.Rodet, Articulation degree as a prosodic dimension of expressive speech, Speech prosody 2008, Campinas, Brazil, May 6-9, pp. 681-684, 2008.

▪ D.Bitouk, A.Nenkova, R.Verma, Improving emotion recognition using class-level spectral features, Interspeech 2009, 6-10 September, Brighton, UK, pp. 2023-2026, 2009.

▪ D.Bitouk, R.Verma, A.Nenkova, Class-level spectral features for emotion recognition, *Speech Communication*, 52, pp. 613-625, 2010.

▪ C.Blouin, V.Maffiolo, A study on the automatic detection and characterization of emotion in a voice service context, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 469- 472, 2005.

▪ R.S.Bolia, R.E.Slyh, Perception of stress and speaking style for selected elements of the SUSAS database, *Speech Communication*, 40, pp. 493-501, 2003.

▪ M.R.Bonner, Changes in the speech pattern under emotional tension, *J. of Psychology*, 56 (2), pp. 262-273, 1943.

▪ N.Boufaden, P.Dumouchel, Leveraging emotion detection using emotions from yes-no answers, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 241-244, 2008.

▪ P.Boula de Mareüil, P.Célérier, J.Toen, Generation of emotions by a morphing technique in English, French and Spanish, Speech Prosody 2002, Aix-en-Provence, France, April 11-13, 2002.

▪ E.Bozkurt, E.Erzin, C.E.Erdem, A.T.Erdem, Improving automatic recognition from speech signals, Interspeech 2009, 6-10 September, Brighton UK, pp. 324-327, 2009.

▪ C.Breitenstein, D.Van Lancker, I.Daum, The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample, *Cognition and Emotion*, 15 (1), pp. 57-79, 2001.

▪ M.Brenner, T.Shipp, Voice stress analysis, NASA report N°88-23395, 1987.

▪ M.Brenner, E.T.Doherty, T.Shipp, Speech measures indicating workload demand, *Aviation Space and Environ. Med.*, 65, pp. 21-26, 1994.

▪ M.Bulut, C.Busso, S.Yildirim, A.Kazemzadeh, C.M.Lee, S.Lee, S.Narayanan, Investigating the role of phoneme-level modifications in emotional speech resynthesis, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 801-804, 2005.

▪ F.Burkhardt, T.Polzehl, J.Stegmann, F.Metze, Detecting real life anger, IEEE ICASSP 2009, April 19-24, Taipei, Taiwan, pp. 4761-4764, 2009.

▪ C.Busso, S.Lee, S.S.Narayanan, Using neutral speech models for emotional speech analysis, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2225-2228, 2007.

▪ J.P.Cabral, L.C.Oliveira, Emovoice: a system to generate emotions in speech, Interspeech 2006 – ICSLP, September 17-21, Pittsburgh, USA, pp. 1798-1801, 2006.

▪ N.Campbell, Databases of emotional speech, ITRW on Speech and Emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 34-38, 2000.

▪ N.Campbell, Perception of affect in speech – towards an automatic processing of paralinguistic information in spoken conversation, Interspeech 2004 – ICSLP 8th Int. Conf. on spoken Language processing, ICC Jeju, Jeju Island, Korea, October 4-8, pp. 881-884, 2004.

▪ N.Chateau, V.Maffiolo, C.Blouin, Analysis of emotional speech in voice mail messages : the influence of speakers' gender, Interspeech 2004, ICSLP, 8th Int. conf. on spoken language processing, ICC Jeju, Jeju Island, Korea, October 4-8, pp. 885-888, 2004.

▪ H.S.Cheang, M.D.Pell, The sound of sarcasm, *Speech Communication*, 50, pp. 366-381, 2008.

▪ H-Y.Cho, K.Yao, T-W.Lee, Emotion verification for emotion detection and unknown emotion rejection, 8th Int. Conf. on Spoken Language Processing, Jeju Island, Korea, October 4-8, pp. 1345-1348, 2004.

▪ Z-J.Chuang, C-H.Wu, Emotion recognition from textual input using an emotional semantic network, 7th Int. Conf. on Spoken Language Processing (ICSLP 2002), Denver, Colorado, USA, September 16-20, pp. 2033-2036, 2002.

▪ S-J.Chung, Perception de l'émotion: les indices prosodiques et la durée minimale, XXIIèmes Journées d'Etudes sur la Parole, Martigny, France, 15-19 Juin, pp. 159-162, 1998.

▪ C.Clavel, I.Vasilescu, L.Devillers, T.Ehrette, G.Richard, Fear-type of the SAFE corpus : annotation issues, Proc of LREC 2006, Genoa, Italy, pp. 1099-1104, 2006.

▪ C.Clavel, I.Vasilescu, L.Devillers, G.Richard, T.Ehrette, Fear-type emotion recognition for future audio-based surveillance systems, Speech Communication, 50, pp. 487-503, 2008.

▪ N.D.Cook, T.Fujisawa, K.Takami, Evaluation of the affect of speech intonation using a model of the perception of interval dissonance and harmonic tension, Eurospeech 2003, Geneva, Switzerland, pp. 105-108, 2003.

▪ R.R.Cornelius, Theoretical approaches to emotion, ITRW on Speech and Emotion, Newcastle, Northern Ireland, UH, September 5-7, pp. 3-10, 2000.

▪ L.Cosmides, Invariances in the acoustic expression of emotion during speech, *J. of Exp. Psychology*, 9 (6), pp. 864-881, 1983.

▪ R.Cowie, E.Douglas-Cowie, Automatic statistical analysis of the signal and prosodic signs of emotion in speech, 4th ICSLP, Philadelphia, USA, October 3-6, pp. 1989-1992, 1996.

▪ R.Cowie, E.Douglas-Cowie, S.Savvidou, E.McMahon, M.Sawey, M.Schröder, Feeltrace: an instrument for recording perceived emotion in real time, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 19-24, 2000.

▪ R.Cowie, Describing the emotional states expressed in speech, ITRW on Speech and Emotion, Newcastle, Northern Ireland, UH, September 5-7, pp. 11-18, 2000.

▪ R.Cowie, E.Douglas-Cowie, N.Tsaptsoulis, W.Fellenz, J.G.Taylor, Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine*, 18 (1), pp. 32-80, January 2001

▪ R.Cowie, R.R.Cornelius, Describing the emotional states that are expressed in speech, *Speech Communication*, 40, pp. 5-32, 2003.

▪ K.E.Cummings, M.A.Clements, Analysis of the glottal excitation of emotionally styled and stressed speech, *J. Acoust. Soc. Am.*, 98 (1), pp. 88-98, 1995.

▪ F.Dellaert, T.Polzin, A.Waibel, Recognizing emotion in speech, 4th Int. Conf. on Spoken Language Processing (ICSLP 96), Philadelphia, USA, October 3-6, 1996.

▪ G.Demenko, Voice stress extraction, Speech Prosody 2008, Campinas, Brazil, May 6-9, pp. 53-56, 2008.

▪ L.Devillers, I.Vasilescu, Prosodic cues for emotion characterization in real-life spoken dialogs, Eurospeech 2003, Geneva, Switzerland, pp. 189- 192, 2003.

▪ L.Devillers, I.Vasilescu, L.Vidrascu, F_0 and pause features analysis for anger and fear detection in real-life spoken dialogs, Speech Prosody 2004, Nara, Japan, march 23-26, 2004.

▪ L.Devillers, I.Vasilescu, Détection des émotions à partir d'indices lexicaux, dialogiques et prosodiques dans le dialogue oral, Journées d'Etudes sur la Parole, 19-22 Avril, Fès, Maroc, 2004.

▪ L.Devillers, L.Vidrascu, Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs, Interspeech 2006, ICSLP, September 17-21, Pittsburgh, USA, pp. 801-804, 2006.

▪ E.Douglas-Cowie, N.Campbell, R.Cowie, P.Roach, Emotional speech: towards a new generation of databases, *Speech Communication*, 40, pp. 33-60, 2003.

▪ E.Douglas-Cowie, L.Devillers, J-C.Martin, R.Cowie, S.Savvidou, S.Abrilian, C.Cox, Multimodal databases of everyday emotion: facing up to complexity, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 813-816, 2005.

▪ A.Drahota, A.Costall, V.Reddy, The vocal communication of different kinds of smile, *Speech Communication*, 50, pp. 278-287, 2008.

▪ P.Dumouchel, N.Dehak, Y.Attabi, R.Dehak, N.Boufaden, Cepstral and long-term features for emotion recognition, Interspeech 2009, 6-10 September, Brighton UK, pp. 344-347, 2009.

▪ C.Drioli, G.Tisato, P.Cosi, F.Tesser, Emotions and voice quality: experiments with sinusoidal modeling, Voqual'03, Geneva, Switzerland, August 27-29, pp. 127-132, 2003.

▪ P.Dumouchel, N.Dehak, Y.Attabi, R.Dehak, N.Boufaden, Cepstral and long-term features for emotion recognition, Interspeech 2009, 6-10 September, Brighton UK, pp. 344-347, 2009.

▪ T.Ehrette, N.Chateau, C.d'Alessandro, V.Maffiolo, Prosodic parameters of perceived emotions in vocal server voices, Speech Prosody 2002, Aix-en Provence, France, April 11-13, pp. 423-426, 2002.

▪ C.Émond, Une analyse prosodique de la parole souriante: étude préliminaire, Actes des XXVIes JEP, Dinard, France, pp. 147-150, 2006.

▪ D.Erickson, A.Abramson, K.Maekawa, T.Kaburagi, Articulatory characteristics of emotional utterances in spoken English, 6th Int. Conf on Spoken Language Processing (ICSLP 2000), Beijing, China, October 16-20, pp. 365-368, 2000.

▪ D.Erickson, C.Menezes, A.Fujino, Some articulatory measurements of real sadness, Interspeech 2004, 8th ICSLP, ICC Jeju, Jeju Island, Korea, October 4-8, pp. 1825-1828, 2004.

▪ D.Erickson, T.Shochi, C.Menezes, H.Kawahara, K-I.Sakakibara, Some non-F_0 cues to emotional speech: an experiment with morphing, Speech Prosody 2008, Campinas, Brazil, May 6-9, pp. 677-680, 2008.

▪ G.Fairbanks, W.Pronovost, Vocal pitch during simulated emotion, *Science*, 88 (2286), pp. 382-383, 1938.

▪ R.Fernandez, R.W.Picard, Classical and novel discriminant features for affect recognition from speech, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 473-476, 2005.

▪ C.Fichtel, K.Hammerschmidt, U.Jürgens, On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the Squirrel monkey, *Behaviour*, 138 (1), pp. 97-116, 2001.

▪ T.Fujisawa, N.D.Cook, Identifying emotion in speech prosody using acoustical cues of harmony, Interspeech 2004 – ICSLP, Jeju Island, Korea, October 4-8, pp. 1333-1336, 2004.

▪ C.Gendrot, Ouverture de la glotte, Fo, intensité et simulations émotionnelles: le cas de la joie, la colère, la surprise, la tristesse et la neutralité, XXIVèmes JEP, Nancy, France, Juin 24-27, pp. 49-52, 2002.

▪ A.Ghio, C.André, B.Teston, C.Cavé, Perceval : une station automatisée de tests de perception et d'évaluation auditive et visuelle, Travaux interdisciplinaires du laboratoire parole et langage, 22, pp. 115-133, 2003.

▪ A.M.Goberman, S.Hughes, T.Haydock, Acoustic characteristics of public speaking: anxiety and practice effects, *Speech Communication*, 53, pp. 867-876, 2011.

▪ M.Goudbeek, J.P.Goldman, K.R.Scherer, Emotions and articulatory precision, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 317-317, 2008.

▪ M.Goudbeek, J.P.Goldman, K.R.Scherer, Emotion dimensions and formant position, Interspeech 2009, 6-10 September, Brighton, UK, pp. 1575-1578, 2009.

▪ M.Goudbeek, M.Broersma, Language specific effects of emotion on phoneme duration, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 2026-2029, 2010.

▪ B.Granström, D.House, Audiovisual representation of prosody in expressive speech communication, *Speech Communication*, 46, pp. 473-484, 2005.

▪ I.Grichkovtsova, A.Lacheret, M.Morel, The role of intonation and voice quality in the affective speech perception, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2245-2248, 2007.

▪ I.Grichkovtsova, M.Morel, A.Lacheret, Identification des émotions en voix naturelle et synthétique : paradigme d'ancrage, JEP 2008, 9-13 Juin, Avignon, France, 2008.

▪ G.R.Griffin, C.E.Williams, The effects of different levels of task complexity on three vocal measures, *Aviation Space and Environ. Med*., 58 (12), pp. 1165-1170, 1987.

▪ M.Grimm, K.Kroschel, E.Mower, S.Narayanan, Primitives-based evaluation and estimation of emotions in speech, *Speech Communication*, 49, pp. 787-800, 2007.

▪ T.Habermas, M.Meier, B.Mukhtar, Are specific emotions narrated differently ?, *Emotion*, 9 (6), pp. 751-762, 2009.

▪ K.Hammerschmidt, U.Jürgens, Acoustical correlates of affective prosody, *J. of Voice*, 21 (5), pp. 531-540, 2007.

▪ A.Hassan, R.I.Damper, Emotion recognition from speech using extended feature selection and a simple classifier, Interspeech 2009, 6-10 September, Brighton, UK, pp. 2043-2046, 2009.

▪ L.He, M.Lech, N.Allen, On the importance of glottal flow spectral energy for the recognition of emotions in speech, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 2346-2349, 2010.

▪ M.H.L.Hecker, K.N.Stevens, G.von Bismarck, C.E.Williams, Manifestations of task-induced stress in the acoustic speech signal, *J. Acoust. Soc. Am*, 44 (4), pp. 993-1001, 1968.

▪ K.Hirose, N.Minematsu, H.Kawanami, Analytical and perceptual study on the role of acoustic features in realizing emotional speech, 6th Int. Conf. on Spoken Language Processing (ICSLP 2000), Beijing, China, October 16-20, pp. 369-372, 2000.

▪ K.Hirose, K.Sato, N.Minematsu, Improvement in corpus-based generation of F_0 contours using generation process model for emotional speech synthesis, Interspeech 2004 – ICSLP, 8th Int. Conf. on Spoken Language Processing, ICC Jeju, Jeju Island, Korea, October 4-8, pp. 1349-1352, 2004.

▪ M.M.Homayounpour, G.Chollet, J.Vaissière, Etude des variabilités intra- et inter-locuteurs provoquées par l'émotion et l'imitation pour les systèmes de vérification du locuteur, XXèmes JEP, Trégastel, France, Juin 1-3, pp. 269-274, 1994.

▪ V.Hozjan, Z.Kačič, Improved emotion recognition with large set of statistical features, Eurospeech 2003, Geneva, Switzerland, pp. 133-136, 2003.

▪ H.Hu, M-X.Xu, W.Wu, Fusion of global statistical spectral features for speech emotion recognition, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2269-2272, 2007.

▪ C-F.Huang, M.Akagi, A three – layered model for expressive speech perception, *Speech Communication*, 50, pp. 810-828, 2008.

▪ D.Hübner, B.Vlasenko, T.Grosser, A.Wendemuth, Determinig optimal feautures for emotion recognition from speech by applying an evolutionary algorithm, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 2358-2361, 2010.

▪ G.Hui, C.Shanguang, S.Guangchuan, Emotion classification of infant voice based on features derived from Teager Energy Operator, Proc. IEEE 2008 Congress on Image and Signal processing, May 27-30, Sanya, China, pp. 333-337, 2008.

▪ Z.Ianoglu, S.Young, Data-driven emotion conversion in spoken English, *Speech Communication*, 51, pp. 268-283, 2009.

▪ A.I.Iliev, M.Scordilis, Spoken emotion recognition using glottal symmetry, EURASIP *Journal on Advances in Signal Processing*, special issue on emotion and mental state recognition from speech, Vol. 2011, Article ID 624575, 11 pages, 2011.

▪ I.Iriondo, R.Guaus, A.Rodriguez, P.Lazaro, N.Montoya, J.M.Blanco, D.Bernadas, J.M.Olivier, D.Tena, L.Longhi, Validation of an acoustical modeling of emotional expression in Spanish using speech sunthesis techniques, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 161-166, 2000.

▪ I.Iriondo, S.Planet, J-C.Socoró, F.Alias, Objective and subjective evaluation of an expressive speech corpus, ITRW on nonlinear speech precessing (NOLISP 07), Paris, France, May 22-25, pp. 15-18, 2007.

▪ C.T.Ishi, H.Ishiguro, N.Hagita, Evaluation of prosodic and voice quality features on automatic extraction of paralinguistic information, Proc. IEEE Int. conf. on intelligent robots and systems, October 9-15, Beijing, China, pp. 374-379, 2006.

▪ C.T.Ishi, H.Ishiguro, N.Hagita, Automatic extraction of paralinguistic information using prosodic features related to Fo, duration and voice quality, *Speech Communication*, 50, pp. 531-543, 2008.

▪ D-N.Jiang, L-H.Cai, Classifying emotion in Chinese speech by decomposing prosodic features, 8th Int. Conf. on Spoken Language Processing, Jeju Island, Korea, October 4-8, pp. 1325-1328, 2004.

▪ T.Johnstone, C.M.ran Reekum, K.Hird, K.Kirsner, K.R.Scherer, Affective speech elicited with a computer game, *Emotion*, 5 (4), pp. 513-518, 2005.

▪ D.Kahneman, Pupillary, heart rate, and skin resistance changes during a mental task, *J. Experimental Psychology*, 79 (1), pp. 164-167, 1969.

▪ N.Kamaruddin, A.Wahab, CMAC for speech emotion profiling, Interspeech 2009, 6-10 September, Brighton, UK, pp. 1991-1994, 2009.

▪ H.Kawahara, H.Matsui, Auditory morphing based of an elastic perceptual distance metric in an interference-free time-frequency representation, IEEE Proc. ICASSP 2003, April 6-10, Hong-Kong, China, pp. 256-259, 2003.

▪ H.Kawatsu, S.Ohno, An analysis of individual differences in the Fo contour and the duration of anger utterances at several degrees, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2213-2216, 2007.

▪ R.Kehrein, The prosody of authentic emotions, Speech prosody 2002, Aix-en Provence, France, April 11-13, 2002.

▪ M.Kienast, W.F.Sendlmeier, Acoustical analysis of spectral and temporal changes in emotional speech, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 92-97, 2000.

▪ S-J.Kim, K-K.Kim, M.Hahn, Study on emotional speech features in Korean with its application to voice contour conversion, Interspeech 2004 – ICSLP, 8th Int. Conf. on Spoken Language Processing, ICC Jeju, Jeju Island, Korea, October 4-8, pp. 1849-1852, 2004.

▪ J.Kim, E.André, M.Rehm, T.Vogt, J.Wagner, Integrating information from speech and physiological signals to achieve emotional sensitivity, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 809-812, 2005.

▪ J.Kim, S.Lee, S.Narayanan, A detailed study of word-position effects on emotion expression in speech, Interspeech 2009, 6-10 September, Brighton, UK, pp.1987-1990, 2009.

▪ G.Klasmeyer, An automatic description tool for time contours and long-term average voice features in large emotional speech databases, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 66-71, 2000.

▪ M.Kobayashi, S.Wada, Time-frequency approach for analysis and synthesis of particular emotional voice, IEEE 2005 Int. Symp. on Signal Processing and Information Technology, December 21, pp. 51-56, 2005.

▪ M.Kockmann, L.Burget, J.Černocký, Brno university of technology system for Interspeech 2009 emotion challenge, Interspeech 2009, 6-10 September, Brighton, UK, pp. 348-350, 2009.

▪ K.Koike, H.Suzuki, H.Saito, Prosodic parameters in emotional speech, 5th Int. Conf. on Spoken Language Processing (ICSLP 98), Sydney, Australia, November 30 – December 4, paper 0996, 1998.

▪ E.Krahmer, M.Swerts, On the role of acting skills for the collection of simulated emotional speech, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 261-264, 2008.

▪ R.Kumar, C.P.Rosé, D.J.Litman, Identification of confusion and surprise in spoken dialogs using prosodic features, Interspeech 2006, September 17-21, Pittsburgh, USA, pp. 1842-1845, 2006.

▪ O-W.Kwon, K.Chan, J.Hao, T-W.Lee, Emotion recognition by speech signals, Eurospeech 2003, Geneva, Switzerland, pp. 125-128, 2003.

▪ D.R.Ladd, K.E.A.Silverman, F.Tolkmitt, G.Bergman, K.R.Scherer, Evidence for the independant function of intonation contour type, voice quality, and F_0 range in signaling speaker affect, *J. Acoust. Soc. Am*, 78 (2), pp. 435-444, 1985.

▪ P.J.Lang, M.M.Bradley, B.N.Cuthbert, International Affective Picture System (IAPS) : technical manual and affective ratings, NIMH Center for the study of emotion and attention, University of Florida, USA, 1997.

▪ A.Laukkanen, E.Vilkman, P.Alku, H.Oksanen, Physical variations related to stress and emotional state: a preliminary study, *J. of Phonetics*, 24, pp. 313-335, 1996.

▪ P.N.Le, E.Ambikairajah, J.Epps, V.Sethu, E.H.C.Choi, Investigation of spectral centroid features for cognitive load classification, Speech Communication, 53, pp. 540-551, 2011.

▪ C-C.Lee, E.Mower, C.Busso, S.Lee, S.Narayanan, Emotion recognition using a hierarchical binary decision tree approach, Interspeech 2009, 6-10 September, Brighton UK, pp. 320-323, 2009.

▪ C.M.Lee, S.Yildirim, M.Bulut, A.Kazemzadeh, C.Busso, Z.Deng, S.Lee, S.Narayanan, Emotion recognition based on phoneme classes, 8th Int. Conf. on Spoken Language Processing, Jeju Island, Korea, October 4-8, pp. 889-892, 2004.

▪ C.M.Lee, S.Narayanan, Emotion recognition using a data-driven fuzzy inference system, Eurospeech 2003, Geneva, Switzerland, pp. 157-160, 2003.

▪ C.M.Lee, S.S.Narayanan, Toward detecting emotions in spoken dialogs, *IEEE Trans. on Speech and Audio Processing*, 13 (2), pp. 293-303, 2005.

▪ S.Lee, S.Yildirim, A.Kazemzadeh, S.Narayanan, An articulatory study of emotional speech production, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 497-500, 2005.

▪ L.Leinonen, T.Hiltunen, I.Linnankoski, M-L.Laakso, Expression of emotional – motivational connotations with a one - word utterance, *J. Acoust. Soc. Am*., 102 (3), pp. 1853-1863, 1997.

▪ T.Li, J.Cheolwoo, Parameter changes across different emotions in human speech, Proc. IEEE Int. Conf on information acquisition, June 27 – July 3, Hong Kong and Macau, China, pp. 91-94, 2005.

▪ P.Lieberman, Perturbations in vocal pitch, *J.Acoust. Soc. Am*., 33 (5), pp. 597-603, 1961.

▪ P.Lieberman, S.B.Michaels, Some aspects of fundamental frequency and enveloppe amplitude as related to the emotional content of speech, *J. Acoust. Soc. Am*, 34 (7), pp. 922-927, 1962.

▪ P.Lieberman, A.Protopapas, B.G.Kanki, Speech production and cognitive deficits on Mt. Everest, *Aviation Space and Environ. Med*., 66, pp. 857-864, 1995.

▪ I.Linnankoski, L.Leinonen, M.Vihla, M-L.Laakso, S.Carlson, Conveyance of emotional connotations by a single word in English, *Speech Communication*, 45, pp. 27-39, 2005.

▪ J.Liscombe, J.Venditti, J.Hirschberg, Classifying subject ratings of emotional speech using acoustic features, Eurospeech 2003, Geneva, Switzerland, pp. 725-727, 2003.

▪ D.J.Litman, K.Forbes-Riley, Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors, *Speech Communication*, 48, pp. 559-590, 2006.

▪ I.Lopez-Moreno, C.Ortego-Resa, J.Gonzalez-Rodriguez, D.Ramos, Speaker dependant emotion recognition using prosodic supervectors, Interspeech 2009, 6-10 September, Brighton, UK, pp.1971-1974, 2009.

▪ I.Luengo, E.Navas, I.Hernáez, Combining spectral and prosodic information for emotion recognition in the Interspeech 2009 emotion challenge, Interspeech 2009, 6-10 September, Brighton, UK, pp. 332-335, 2009.

▪ M.Lugger, B.Yang, On the relevance of high-level features for speaker independent emotion recognition of spontaneous speech, Interspeech 2009, 6-10 September, Brighton UK, pp. 1995-1998, 2009.

▪ S.McGilloway, R.Cowie, E.Douglas-Cowie, Approaching automatic recognition of emotion from voice: a rough benchmark, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 207-212, 2000.

▪ E.Mendoza, G.Carballo, Acoustic analysis of induced vocal stress by means of cognitive workload tasks, *J. of Voice*, 12 (3), pp. 263-273, 1998.

▪ E.Mendoza, G.Carballo, Vocal tremor and psychological stress, *J. of Voice*, 13 (1), pp. 105-112, 1999.

▪ C.Menezes, B.Pardo, D.Erickson, O.Fujimura, Changes in syllable magnitude and timing due to repeated correction, *Speech Communication*, 40, pp. 71-85, 2003.

▪ C.Menezes, K.Maekawa, Paralinguistic effects on voice quality: a study in Japanese, Speech prosody 2006, Dresden, Germany, May 2-5, PS6-04 0049, 2006.

▪ D.Morrison, R.Wang, L.C.de Silva, Ensemble methods for spoken emotion recognition in call-centres, *Speech Communication*, 49, pp. 98-112, 2007.

▪ J.D.Mosko, K.N.Stevens, G.R.Griffin, Interactive voice technology: variations in the vocal utterances of speakers performing a stress-inducing task, Naval Aerospace Medical Research Laboratory report N°1300, Florida, USA, 1983.

▪ S.J.L.Mozziconacci, D.J.Hermes, Variations temporelles communiquant l'émotion dans la parole, XXIIIèmes JEP, Aussois, France, Juin 19-23, pp. 145-148, 2000.

▪ S.J.L.Mozziconacci, The expression of emotion considered in the framework of an intonational model, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 45-52, 2000.

▪ S.J.L.Mozziconacci, D.J.Hermes, Expression of emotion and attitude through temporal speech variations, 6th Int. Conf. on Spoken Language Processing (ICSLP 2000), Beijing, China, October 16-20, pp. 373-378, 2000.

▪ I.R.Murray, J.L.Arnott, Toward the simulation of emotion in synthetic speech : a review of the literature on human vocal emotion, *J. Acoust. Soc. Am*, 93 (2), pp. 1097-1108, 1993.

▪I.R.Murray, J.L.Arnott, E.A.Rohwer, Emotional stress in synthetic speech : progress and future directions, *Speech Communication*, 20, pp. 85-91, 1996.

▪ Y.Nagasaki, T.Komatsu, Can people perceive different emotions from a non-emotional voice by modifying its F_0 and duration ?, Speech Prosody 2004, Nara, Japan, March 23-26, pp. 667-670, 2004

▪ D.Neiberg, K.Elenius, K.Laskowski, Emotion recognition in spontaneous speech using GMMs, Interspeech 2006, ICSLP, September 17-21, Pittsburgh, USA, pp. 809-812, 2006.

▪ D.Neiberg, K.Elenius, Automatic recognition of anger in spontaneous speech, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 2255-2275, 2008.

▪ A.Novak, J.Vokral, Emotions in the sight of long time averaged spectrum (LTAS) and three dimensional analysis of periodicity, *Folia Phoniatr*., 45(4), pp. 198-203, 1993.

▪ T.L.Nwe, S.W.Foo, L.C.de Silva, Speech emotion recognition using hidden Markov models, *Speech Communication*, 41, pp. 603-623, 2003.

▪ P-Y.Oudeyer, Novel useful features and algorithms for recognition of emotions in human speech, Speech Prosody 2002, Aix-en-Provence, France, April 11-13, pp. 547-550, 2002.

▪ P-Y.Oudeyer, The synthesis of cartoon emotional speech, Speech Prosody 2002, Aix-en-Provence, France, April 11-13, pp. 551-554, 2002.

▪ C.Ovesdotter Alm, X.Llorà, Evolving emotional prosody, Interspeech 2006 – ICSLP, September 17-21, Pittsburgh, USA, pp. 1826-1829, 2006.

▪ A.Paeschke, W.F.Sendlmeier, Prosodic characteristics of emotional speech: measurements of fundamental frequency movements, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 75-80, 2000.

▪ A.Paeschke, Global trend of fundamental frequency in emotional speech, Speech Prosody 2004, Nara, Japan, March 23-26, pp. 671-674, 2004.

▪ C.Pereira, C.Watson, Some acoustic characteristics of emotion, 5th ICSLP 98, Sidney, Australia, November 30 – December 4, pp. 927-929, 1998.

▪ C.Pereira, Dimensions of emotional meaning in speech, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 25-28, 2000.

▪ V.A.Petrushin, Voice messaging system that organizes voice messages based on detected emotion, United States Patent N° US 6,697,457 B2, Feb. 24, 2004.

▪ S.Planet, I.Iriondo, J-C.Socoró, C.Monzo, J.Adell, GTM-URL contribution to the Interspeech 2009 emotion challenge, Interspeech 2009, 6-10 September, Brighton UK, pp. 316-319, 2009.

▪ T.Polzehl, S.Sundaram, H.Ketabdar, M.Wagner, F.Metze, Emotion classification in children's speech using fusion of acoustic and linguistic features, Interspeech 2009, 6-10 September, Brighton UK, pp. 340-343, 2009.

▪ T.S.Polzin, A.Waibel, Emotion – sensitive Human – computer interfaces, ITRW on speech and emotion, Newcastle, Northern Ireland, UK, September 5-7, pp. 201-206, 2000.

▪ S.R.M.Prasanna, D.Govind, Analysis of excitation source information in emotional speech, Interspeech 2010, September 26-30, Makuhari, Japan, pp. 781-784, 2010.

▪ A.A.Razak, M.I.Z.Abidin, R.Komiya, A preliminaty speech analysis for recognizing emotion, IEEE Student conference on research and development (SCOReD) 2003 Proc., Putrajaya, Malaysia, pp. 49-54, 2003.

▪ F.Ringeval, M.Chetouani, A vowel based approach for acted emotion recognition, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 2763-2766, 2008.

▪ E.Rodero, Intonation and emotion: influence of pitch levels and contour type on creating emotions, *J. of Voice*, 25(1), pp. e25-e34, 2010.

▪ L.J.M.Rothkrantz, P.Wiggers, J.W.A.van Wees, R.J.van Vark, Voice stress analysis, Proc. of Text Speech and Dialogues TSD 2004, pp. 449-456, 2004.

▪ L.Rubenstein, Electro-acoustical measurement of vocal responses to limited stress, *Behav. Res. & Therapy*, 4, pp. 135-138, 1966.

▪ G.Salvi, F.Tesser, E.Zovato, P.Cosi, Cluster analysis of differential spectral envelopes on emotional speech, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 322-325, 2010.

▪ R.Sarikaya, J.N.Gowdy, Subband based classification of speech under stress, Proc. of IEEE ASSP, May 12-15, Seattle, USA, pp. 569-572, 1998

▪ H.Sato, Y.Mitsukura, M.Fukumi, N.Akamatsu, Emotional speech classification with prosodic parameters by using neural networks, 7th Australian and NewZealand Intelligent Information Systems Conference, Perth, Western Australia, 18-21 November, pp. 395-398, 2001.

▪ K.R.Scherer, Vocal indicators of stress, in *Speech evaluation in psychiatry*, pp.171-187, Grune & Stratton, 1981.

▪ K.R.Scherer, Speech and emotional states, in *Speech evaluation in psychiatry*, pp. 189-221, Grune & Stratton, 1981.

▪ K.R.Scherer, T.Johnstone, J.Sangsue, L'état émotionnel du locuteur: facteur négligé mais non négligeable pour la technologie de la parole, XXIIèmes JEP, Martigny, France, Juin 15-19, pp. 249-257, 1998.

▪ K.R.Scherer, A cross-cultural investigation of emotion inferences from voice and speech: implications for speech technology, 6th Int. Conf. on Spoken Language Processing (ICSLP 2000), Beijing, China, October 16-20, pp. 379-382, 2000.

▪ K.R.Scherer, D.Grandjean, T.Johnstone, G.Klasmeyer, T.Banziger, Acoustic correlates of task load and stress, 7th Int. Conf. on Spoken Language Processing (ICSLP2002), Denver, USA, September 16-20, pp. 2017-2020, 2002.

▪ K.R.Scherer, Vocal communication of emotion: a review of research paradigms, *Speech Communication*, 40, pp. 227-256, 2003.

▪ M.Schröder, Experimental study of affect bursts, ITRW on speech and emotion, Northern Ireland, UK, September 5-7, pp. 132-137, 2000.

▪ M.Schröder, R.Cowie, E.Douglas-Cowie, M.Westerdijk, S.Gielen, Acoustic correlates of emotion dimensions in view of speech synthesis, Eurospeech 2001, Scandinavia, pp. 87-90, 2001.

▪ M.Schröder, Experimental study of affect bursts, *Speech Communication*, 40, pp. 99-116, 2003.

▪ B.Schuller, R.Müller, M.Lang, G.Rigoll, Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 805-808, 2005.

▪ B.Schuller, A.Batliner, D.Seppi, S.Steidl, J.Wagner, L.Devillers, N.Amir, L.Kessous, V.Aharonson, The relevance of feature type for the automatic classification of emotional user

states: low level descriptors and functional, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2253-2256, 2007.

▪ B.Schuller, S.Steidl, A.Batliner, The Interspeech 2009 Emotion Challenge, Interspeech 2009, 6-10 September, Brighton, UK, pp. 312-315, 2009.

▪ D.Seppi, A.Batliner, B.Schuller, S.Steidl, T.Vogt, J.Wagner, L.Devillers, L.Vidrascu, N.Amir, V.Aharonson, Patterns, prototypes, performance: classifying emotional user states, Interspeech 2008, September 22-26, Brisbane Australia, pp. 601-604, 2008.

▪ V.Sethu, E.Ambikairajah, J.Epps, Group delay features for emotion detection, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2273-2276, 2007.

▪ V.Sethu, E.Ambikairajah, J.Epps, Phonetic and speaker variations in automatic emotion classification, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 617-620, 2008.

▪ V.Sethu, E.Ambikairajah, J.Epps, Pitch contour parametrisation on linear stylisation for emotion recognition, Interspeech 2009, 6-10 September, Brighton, UK, pp. 2011-2014, 2009.

▪ A.Shaukat, K.Chen, Towards automatic emotional state categorization from speech signals, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 2771-2774, 2008.

▪ A.W.Siegman, S.Boyle, Voices of fear and anxiety and sadness and depression: the effects of speech rate and loudness on fear and anxiety and sadness and depression, *J. of Abnormal Psychology*, 102 (3), pp. 430-437, 1993.

▪ E.R.Simon-Thomas, D.J.Keltner, D.Sauter, L.Sinicropi-Yao, A.Abramson, The voice conveys specific emotions: evidence from vocal burst displays, *Emotion*, 9 (6), pp. 838-846, 2009.

▪ P.V.Simonov, M.V.Frolov, V.L.Taubkin, Use of the invariant method of speech analysis to discern the emotional state of announcers, *Aviation Space and Environ. Med*., 46 (8), pp. 1016-1016, 1975.

▪ T.Seppänen, E.Väyrynen, J.Toivanen, Prosody-based classification of emotions in spoken Finnish, Eurospeech 2003, Geneva, Switzerland, pp. 717-720, 2003.

▪ M.Shami, W.Verhelst, An evaluation of the robustness of existing supervised machine learning approaches to the classification of emotions in speech, *Speech Communication*, 49, pp. 201-212, 2007.

▪ K.Ślot, J.Cichosz, L.Bronakowski, Application of voiced-speech variability descriptors to emotion recognition, Proc. 2009 IEEE Symp. On computational intelligence in security and defense appl (CISDA 2009), July 8-10, Ottawa, pp.1-5, 2009.

▪ K.Soltani, R.N.Ainon, Speech emotion detection based on neural networks, 9th Int. Symp. on signal processing and its application (ISSPA), Feb. 12-15, Sharjah, United Arab Emirate, pp. 1-3, 2007.

▪ J.Soltis, K.A.Leighty, C.M.Wesolek, A.Savage, The expression of affect in African elephant (Loxodonta Africana) rumble vocalizations, *J. of Comparative Psychology*, 123 (2), pp. 222-225, 2009.

▪ R.Stibbard, Automated extraction of ToBi annotation data from the reading/leeds emotional speech corpus, ITRW on Speech end Emotion, Northern Ireland, UK, September 5-7, pp. 60-65, 2000.

▪ L.A.Astreeter, N.H.Macdonald, W.Apple, R.M.Krauss, K.M.Galotti, Acoustic and perceptual indicators of emotional stress, *J. Acoust. Soc. Am*., 73 (4), pp. 1354-1360, 1983.

▪ T.Suzuki, M.Ikemoto, T.Sano, T.Kinoshita, Multidimensional features of emotional speech, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 240-240, 2008.

▪ M.Swerts, E.Krahmer, Gender – related differences in the production and perception of emotion, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 334-337, 2008.

▪ S.Takeda, Y.Yasuda, R.Isobe, S.Kiryu, M.Tsuru, Analysis of voice-quality features of speech that expresses “anger”, “joy”, and “sadness” uttered by radio actors and actresses, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 2114-2117, 2008.

▪ J.Tao, Y.Kang, A.Li, Prosody conversion from neutral speech to emotional speech, *IEEE Trans. on Audio, Speech, and Language Processing*, 14 (4), pp. 1145-1154, 2006.

▪ R.Tato, R.Santos, R.Kompe, J.M.Pardo, Emotional space improves emotion recognition, 7th Int. Conf. on Spoken Language Processing (ICSLP 2002), Denver, Colorado, USA, September 16-20, pp. 2029-2032, 2002.

▪ F.J.Tolkmitt, K.R.Scherer, Effect of experimentally induced stress on vocal parameters, *J. Experimental Psychology: Human Perception and Performance*, 12 (3), pp. 302-313, 1986.

▪ K.P.Truong, D.A.van Leeuwen, Visualizing acoustic similarities between emotions in speech: an acoustic map of emotions, Interspeech 2007, August 27-31, Antwerp, Belgium, pp. 2265-2268, 2007.

▪ J.Trouvain, W.J.Barry, The prosody of excitement in horse race commentaries, ITRW on Speech and Emotion , Newcastle, Northern Ireland, UK, September 5-7, pp. 86-91, 2000.

▪ J.Trouvain, S.Schmidt, M.Schröder, M.Schmitz, W.J.Barry, Modelling personality features by changing prosody in synthetic speech, Speech Prosody 2006, Dresden, Germany, May 2-5, 2006.

▪ K.P.Truong, D.A.van Leeuwen, M.A.Neerincx, F.M.G.de Jong, Arousal and valence prediction in spontaneous emotional speech: felt versus perceived emotion, Interspeech 2009, 6-10 September, Brighton UK, pp. 2027-2030, 2009.

▪ B.Tursky, G.E.Schwartz, A.Crider, Differential patterns of heart rate and skin resistance during a digit-transformation task, *J. Experimental Psychology*, 83 (3), pp. 451-457, 1970.

▪ E.Väyrynen, J.Toivanen, T.Seppänen, Classification of emotion in spoken Finnish using vowel-length segments: inceasing reliability with a fusion technique, Speech Communication, 53, pp. 269-282, 2011.

▪ D.Ververidis, C.Kotropoulos, Automatic speech classification to five emotional states based on gender information, EUSIPCO 2004, pp. 341-344, 2004.

▪ D.Ververidis, C.Kotropoulos, Emotional speech classification using Gaussian mixture models, Proc. IEEE ISCAS, May 23-26, Kobe, Japan, pp. 2871-2874, 2005.

▪ D.Ververidis, C.Kotropoulos, Emotional speech recognition: resources, features, and methods, *Speech Communication*, 48, pp. 1162-1181, 2006.

▪ L.Vidrascu, L.Devillers, Detection of real-life emotions in call centers, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 1841-1844, 2005.

▪ L.Vidrascu, L.Devillers, Détection de 2 à 5 classes d’émotions dans les dialogues enregistrés dans un centre d’appel, Actes des VIIèmes RJC Parole, Paris, 5-6 Juillet, pp. 148-151, 2007.

▪ T.Vogt, E.André, Exploring the benefits of discretization of acoustic features for speech emotion recognition, Interspeech 2009, 6-10 September, Brighton UK, pp. 328-331, 2009.

▪ J.Vroomen, R.Collier, S.Mozziconacci, Duration and intonation in emotional speech, 3rd European Conf. on Speech Communication and Technology, Eurospeech’93, Berlin, Germany, September 19-23, pp. 577-580, 1993.

▪ T.Waaramaa, A-M.Laukkanen, M.Airas, P.Alku, Perception of emotional valences and activity levels from vowel segments of continuous speech, *J. of Voice*, 24(1), pp. 30-38, 2010.

▪ S.P.Whiteside, Simulated emotions: an acoustic study of voice and perturbation measures, 5th ICSLP 98, Sidney, Australia, November 30 – December 4, paper 0153, 1998.

▪ C.E.Williams, K.N.Stevens, Emotions and speech: some acoustical correlates, *J. Acoust. Soc. Am.*, 52 (4), pp. 1238-1250, 1972.

▪ J.Wilting, E.Krahmer, M.Swerts, Real vs acted emotional speech, Interspeech 2006 – ICSLP, September 17-21, Pittsburgh, USA, pp. 805-808, 2006.

▪ M.Wöllmer, F.Eyben, S.Reiter, B.Schuller, C.Cox, E.Douglas-Cowie, R .Cowie, Abandonning emotion classes – towards continuous emotion recognition with modeling of long-range dependencies, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 597-600, 2008.

▪ M.Wöllmer, A.Metallinou, F.Eyben, B.Schuller, S.Narayanan, Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional LSTM modeling, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 2362-2365, 2010.

▪ W.Wu, T.F.Zheng, M-X.Xu, H-J.Bao, Study on speaker verification on emotional speech, Interspeech 2006, September 17-21, Pittsburgh, USA, pp. 2102- 2105, 2006.

▪ D.Wu, T.D.Parsons, S.S.Narayanan, Acoustic feature analysis in speech emotion primitives estimation, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 785-788, 2010.

▪ S.Wu, T.H.Falk, W-Y.Chan, Automatic speech emotion recognition using modulation spectral features, *Speech Communication*, 53, pp. 768-785, 2011.

▪ L.Xu, M.Xu, D.Yang, ANN based decision fusion for speech emotion recognition, Interspeech 2009, 6-10 September, Brighton, UK, pp. 2035-2038, 2009.

▪ S.Yacoub, S.Simske, X.Lin, J.Burns, Recognition of emotions in interactive voice response systems, Eurospeech 2003, Geneva, Switzerland, pp. 729-732, 2003.

▪ H.Yamasaki, Perception des émotions « positives » et « négatives » chez les auditeurs français et japonais à travers le contour de F_0, JEP 2004, 19-22 Avril, Fès, Maroc, 2004.

▪ T.F.Yap, J.Epps, E.Ambikairajah, E.H.C.Choi, An investigation of formant frequencies for cognitive load classification, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 2022-2025, 2010.

▪ T.F.Yap, J.Epps, E.Ambikairajah, E.H.C.Choi, Formant frequencies under cognitive load: effects and classification, *Eurasip J. on Advances in Signal Processing*, special issue on emotional mental state recognition from speech, article ID 219253, 11 pages, 2011.

▪ J.Yuan, L.Shen, F.Chen, The acoustic realization of anger, fear, joy and sadness in Chinese, 7th ICSLP, Denver, USA, September 16-20, pp. 2025-2028, 2002.

▪ T.Zhang, M.Hasegawa-Johnson, S.E.Levinson, Children's emotion recognition in an intelligent tutoring scenario, Interspeech 2004, 8th ICSLP, ICC Jeju, Jeju Island, Korea, October 4-8, pp. 1441-1444, 2004.

▪ B.Zei, M.Archimard, La variabilité du rythme cardiaque et la différentiation prosodique des émotions, XXIIèmes J.E.P, Martigny 15-19 Juin, pp. 167-170, 1998.

▪ Y.Zhou, Y.Sun, J.Li, J.Zhang, Y.Yan, Physiologically-inspired feature extraction for emotion recognition, Interspeech 2009, 6-10 September, Brighton, UK, pp.1975-1978, 2009.

5.2.3. Voix et Fatigue

▪ A.D.Bagnall, The impact of fatigue on the voice, PhD Thesis, School of Psychology, University of South Australia, October 2007.

▪ A.D.Bagnall, J.Dorrian, A.Fletcher, Some vocal consequences of sleep deprivation and the possibility of "fatigue proofing" the voice with voicecraft® voice training, *J.of Voice*, sous presse, 2011

▪ W.A.Bousfield, The influence of fatigue on tremor, *J. Exp. Psychology*, 15 (1), pp. 104-107, 1932.

▪ M.J.Caraty, C.Montacié, Multivariate analysis of vocal fatigue in continuous reading, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 470-473, 2010.

▪ J.Davidović, S. Ćićević, Psycholinguistic meaning of the preformant in the vowel acoustic structure, 5th Int. Cong. of the Int. Soc. Applied Psycholinguistics, June 25-27, Porto, Portugal, pp. 139-142, 1997.

▪ H.P.Greeley, E.Friets, J.P.Wilson, S.Raghavan, J.Picone, J.Berg, Detecting fatigue from voice using speech recognition, IEEE Int. Symp. on Signal Processing and Information Technology, pp. 567-571, 2006.

▪ H.P.Greeley, J.Berg, E.Friets, J.Wilson, Fatigue estimation using voice analysis, *Behavior Research Methods*, 39 (3), pp. 610-619, 2007.

▪ Y.Harrison, J.A.Horne, Sleep deprivation affects speech, *Sleep*, 20(10), pp. 871-877, 1997.

▪ H.S.Hayre, J.C.Holland, Cross-correlation of voice and heart rate as stress measures, *Applied Acoustics*, 13, pp. 57-62, 1980.

▪ J.Krajewski, A.Batliner, M.Golz, Acoustic sleepiness detection: framework and validation of a speech-adapted pattern recognition approach, *Behavior Research Methods*, 41(3), pp. 795-804, 2009.

▪ Q.Ji, Z.Zhu, P.Lan, Real-time nonintrusive monitoring and prediction of driver fatigue, *IEEE Trans. on Vehicular Technology*, 53 (4), pp. 1052-1068, 2004.

▪ T.L.Nwe, H.Li, M.Dong, Analysis and detection of speech under sleep deprivation, Interspeech 2006, Septembre 17-21, Pittsburgh, USA, pp. 1846-1849, 2006.

▪ S.A.Patil, J.H.L.Hansen, Detection of speech under physical stress: model development, sensor selection, and feature fusion, Interspeech 2008, September 22-26, Brisbane, Australia, pp. 817-820, 2008.

▪ K.Shiomi, S.Hirose, Fatigue and drowsiness predictor for pilots and air traffic controllers, 45th ATCA conf., Oct 22-26, Atlantic city, USA, pp. 1-4, 2000.

▪ K.Shiomi, Chaologic index value calculation system, European Patent Application N° EP 1 598 749 A1, December 26, 2003.

▪ K.Shiomi, M.Takaoka, Chaotic voice analysis and model of cerebral dynamics, *Kansei Engineering*, 4 (1), pp. 3-12, 2004.

▪ K.Shiomi, S.Hirose, Apparatus for detecting fatigue and doze by voice, and recording medium, United States Patent N° US 6,876,964 B1, Apr. 5, 2005.

▪ K.Shiomi, K.Sato, M.Sawa, N.Mizukami, A.Suzuki, Experimental results of measuring human fatigue by utilizing uttered voice processing, IEEE Int. Conf. on Systems, Man and Cybernetics, Singapore, pp. 3338-3342, 2008.

▪ K.Shiomi, Voice processing technique for human cerebral activity measurement, Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics, Singapore, P0660, 2008.

▪ K.Shiomi, Chaotic voice analysis method for human performance monitoring, Proc. European Safety and Relaibility Ass. ESREL, September 7-10, Prague, Czech Republic, 2009.

5.2.4. Voix et Autres Facteurs Perturbateurs

▪ M.Alpert, R.L.Kurtzberg, A.J.Friedhoff, Transient voice changes associated with emotional stimuli, *Archives of General Psychiatry*, 8, pp. 362-365, 1963.

▪ A.D.Bell, Jr., W.H.Ford, C.R.McQuiston, Physiological response analysis method and apparatus, United States patent N°3,971,034, July 20, 1976.

▪ S.Benus, F.Enos, J.Hirschberg, E.Shriberg, Pauses in deceptive speech, Speech Prosody 2006, Dresden, Germany, May 2-5, 2006.

▪ J-F.Bonastre, D.Matrouf, C.Fredouille, Augmentation du taux de fausse acceptation par transformation inaudible de la voix des imposteurs, Actes des XXVIes JEP, Dinard, France, pp. 11-14, 2006.

▪ M.Brenner, H.H.Branscomb, G.E.Schwartz, Psychological stress evaluator – two tests of a vocal measure, *Psychophysiology*, 16 (4), pp. 351-357, 1979.

▪ V.L.Cestaro, A test of the computer voice stress analyzer (CVSA) theory of operation, *Polygraph*, 31 (2), pp. 72-86, 2002.

▪ H.Charles, Method for quantifying psychological stress levels using voice pattern samples, United States Patent Application Publication N° US 2005/0131692 A1, Jun. 16, 2005.

▪ C.Chevrie-Muller, N.Seguier, A.Spira, M.Dordain, Recognition of psychiatric disorders from voice quality, *Language and Speech*, 21 (1), pp. 87-111, 1978.

▪ C.Chevrie-Muller, P.Sevestre, N.Seguier, Speech and psychopathology, *Language an Speech*, 28 (1), pp. 57-79, 1985.

▪ R.Cowie, E.Douglas-Cowie, Automatic statistical analysis of the signal and prosodic signs of emotion in speech, 4th ICSLP, Philadelphia, USA, October 3-6, 1996.

▪ J.K.Darby, *Speech Evaluation in Psychiatry*, Grune & Stratton, 1981.

▪ F.Enos, S.Benus, R.L.Cautin, M.Graciarena, J.Hirschberg, E.Shriberg, Personality factors un human deception detection: comparing human to machine performance, Interspeech 2006 - ICSLP, September 17-21, Pittsburgh, USA, pp. 813-816, 2006

▪ A.J.Friedhoff, M.Alpert, R.J.Kurtzberg, An electro-acoustic analysis of the effects of stress on voice, *J. of Neuropsychiatry*, 5, pp. 266-272, 1964.

▪ F.H.Fuller, Method and apparatus for phonation analysis leading to valid truth/lie decisions by vibrato component assessment, United States Patent N° 3,855,418, Dec. 17, 1972.

▪ F.H.Fuller, Method and apparatus for phonation analysis leading to valid truth/lie decisions by spectral energy region comparison, United States patent N° 3,855,417, Dec. 17, 1974.

▪ P.Furness, J.Jessop, O.C.J.Lippold, Long-lasting increases in the tremor of human and muscles following brief, strong effort, *J. Physiol*., 265, pp. 821-831, 1977.

▪ D.Haddad, S.Walter, R.Ratley, M.Smith, Investigation and evaluation of voice stress analysis technology, Doc. N° 193832, Office of Justice programs, National Institute of Justice, Department of Justice, February 13, 2002.

▪ J.H.L.Hansen, G.Zhou, B.L.Pellom, Methods for voice stress analysis and classification, Final Technical Report, Rome, NY, U.S. Air Force Research Laboratory, 1999.

▪ J.Hirschberg, S.Benus, J.M.Brenier, Distinguishing deceptive from non-deceptive speech, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 1833- 1836, 2005.

▪ H.Hollien, J.D.Harnsberger, C.A.Martin, R.Hill, G.A.Alderman, Perceiving the effects of ethanol intoxication on voice, *J. of Voice*, 23(5), pp. 552-559, 2009.

▪ F.Horvath, An experimental comparison of the physiological stress evaluator and the galvanic skin response in detection of deception, *J. Applied Psychology*, 63 (3), pp. 338-344, 1978.

▪ F.Klingholz, R.Penning, E.Liebhardt, Recognition of low-level alcohol intoxication from speech signals, *J. Acoust. Soc. Am*, 84 (3), pp. 929-935, 1988.

▪ A.Liberman, Apparatus and methods for detecting emotions, United States Patent N° 6,638,217 B1, Oct. 28, 2003.

▪ O.Lippold, Physiological tremor, *Scientific American*, 224 (3), pp. 65-73, 1971.

▪ J.K.Y.Ma, R.Hoffmann, Acoustic analysis of intonation in parkinson's disease, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 2586-2589, 2010.

▪ E.Mendoza, G.Carballo, Vocal tremor and psychological stress, *J. of Voice*, 13 (1), pp. 105-112, 1999.

▪ A.Nilsonne, J.Sundberg, S.Ternström, A.Askenfelt, Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression, *J. Acoust. Soc. Am*, 83 (2), pp. 716-728, 1988.

▪ P.O'Neal Nunally, Audio psychological stress indicator alteration method and apparatus, United States Patent N° US 7,191,134,B2, Mar. 13, 2007.

▪ D.B.Pisoni, C.S.Martin, Effects of alcohol on the acoustic-phonetic properties of speech: perceptual and acoustic analyses, *Alcohlism: Clinical and Experimental Research*, 13(4), pp. 577-587, 1989.

▪ D.Rochman, G.M.Diamond, O.Amir, Unresolved anger and sadness: identifying vocal acoustical correlates, *Journal of Counseling Psychology*, 55 (4), pp. 505-517, 2008.

▪ A.Rosenberg, J.Hirschberg, Acoustic/prosodic and lexical correlates of charismatic speech, Interspeech 2005, September 4-8, Lisbon, Portugal, pp. 513-516, 2005.

▪ E.Scheiner, K.Hammerschmidt, U.Jürgens, P.Zwirner, Vocal expression of emotions in normally hearing and hearing-impaired infants, *J. of Voice*, 20 (4), pp. 585-604, 2006.

▪ F.Schiel, C.Heinrich, V.Neumeyer, Rhythm and formant features for automatic alcohol detection, Interspeech 2010, September 26-30, Makuhari, Chiba, Japan, pp. 458-461, 2010.

▪ S.G.Schiflett, G.J.Loikith, Voice stress analysis as a measure of operator workload, Technical Memorandum TM 79-3 SY, Naval Air Test Center, USA, 1980.

▪ G.A.Smith, Voice analysis for the measurement of anxiety, *Br. J. med. Psychol*., 50, pp. 367-373, 1977.

▪ L.A.Streeter, R.M.Krauss, V.Geller, C.Olson, W.Apple, Pitch changes during attempted deception, *J. of Personality and Social Psychology*, 35 (5), pp. 345-350, 1977.

▪ J.D.Williamson, Speech analyzer for analyzing frequency perturbations in a speech pattern to determine the emotional state of a person, United State Patent N° 4,142,067, Feb. 27, 1979.

▪ W.S.Winholtz, L.O.Ramig, Vocal tremor analysis with the vocal demodulator, *J. of Speech and Hearing Research*, 35, pp. 562-573, 1992.

5.2.5. Ouvrages Généraux et Publications Complémentaires

Ces ouvrages et ces publications ne sont pas tous référencés dans le manuscrit. Ils sont mentionnés car leur lecture a contribué à l'étude, qu'il s'agisse de traitement du signal vocal ou de considérations générales sur la parole.

▪ Anthropologie appliquée, Étude de la fatigue en aéronautique, phase III : synthèse des résultats et recommandations, rapport du laboratoire d'anthropologie appliquée, Université Paris V, DOC AA 395/98, Novembre 1998.

▪ J.Benesti, M.M.Sondhi, Y.Huang: Handbook of speech processing, Springer, 2008.

▪ R.Boite, M.Kunt, Traitement automatique de la parole, Presses Polytechniques Romandes, 1987.

▪ Calliope, La Parole et son traitement automatique, Masson et CNET-ENST, 1989.

▪ A.R.Damasio, L'erreur de Descartes, la raison des émotions, Odile Jacob, 1995.

▪ Ch.Darwin, L'expression des émotions chez l'homme et les animaux, trad. D.Férault, Rivages Poche, 2001 (édition originale, 1872).

▪ J.L.Dessalles, Aux origines du langage, une histoire naturelle de la parole, Hermès Lavoisier, 2000.

▪ P.Escudier, J.L.Schwartz, La parole, des modèles cognitifs aux machines communicantes, Hermès, 2000.

▪ J.L.Flanagan, Speech analysis, synthesis and perception, Springer-Verlag, 1972.

▪ A.Giovanni, M.Ouakine, J.M.Triglia, Determination of largest Lyapunov exponents of vocal signal, *J. of Voice*, 13 (3), pp. 341-354, 1999.

▪ H.Herzel, D.Berry, I.R.Titze, M.Saleh, Analysis of vocal disorders with methods from nonlinear dynamics. *J. Speech and Hearing Disorders*, 37, pp. 1008-1019, 1994.

▪ W.Hess, Pitch determination of speech signals, Springer-Verlag, 1983.

▪ H.Kantz, T.Schreiber, Nonlinear time series analysis. Cambridge University Press, 2005.

▪ A.Karpf, La Voix, éd. Autrement, 2008.

▪ M.Lewis, J.M.Haviland-Jones, L.Feldman Barret, Handbook of emotions, 3rd edition, The Guilford Press, 2008.

▪ J.Mariani, Analyse, synthèse et codage de la parole, Hermès Lavoisier, 2002.

▪ J.D.Markel, A.H.Gray Jr., Linear prediction of speech, Springer-Verlag, 1976.

▪ M.T.Rosenstein, J.J.Collins, C.J.De Luca, A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D*, 65, pp. 117-134, 1993.

▪ J.C.Sprott, Chaos and time-series analysis, Oxford University Press, 2006.

▪ Test de Stress, manuel, Eds. Centre de Psychologie Appliquée, Paris, 1975.

TABLE des MATIÈRES

1. Introduction : Cadre de la Recherche…….……2

2. Revue Générale des Connaissances……..……7

2.1 Analyse de la Voix dans le Secteur des Transports……..…...8

2.2 Voix et Émotion……...…….15
2.2.1 Émotions Simulées ou Induites……...…..17
2.2.1.1 Émotions Simulées par des Acteurs……....18
2.2.1.1.1 Analyse Acoustique…….18
2.2.1.1.2 Reconnaissance Automatique……..21
2.2.1.2 Émotions Induites par le Contexte Expérimental……....27
2.2.1.3 Perception des Émotions ……....30
2.2.1.4 Émotions Simulées par Synthèse Vocale ……...32
2.2.2 Émotions Réelles……...35
2.2.3 Stress Cognitif et Charge de Travail…….39

2.3 Voix et Fatigue……..41

2.4 Voix et Autres Facteurs Perturbateurs……...43
2.4.1 Voix et Mensonge……...……...…43
2.4.2 Voix et Pathologies Non Vocales……..44
2.4.3 Voix et Alcoolémie……...45

2.5 Synthèse de la Revue Générale des Connaissances……..…46

3. Contribution de l'Analyse Acoustique de la Voix au Contexte Aéronautique et au Domaine de la Psychologie…….……48

3.1 Émotion…….……49
3.1.1 Enregistreur Phonique des Conversations en Cockpit……..…49
3.1.2 Vol Long-Courrier…….56
3.1.3 Simulateur de Vol……..59

3.2 Fatigue…….……..62
3.2.1 Vols Court-Courriers Consécutifs……63
3.2.2 Conducteur Automobile……67
3.2.3 Inertie du Sommeil……...……70
3.2.4 Fatigue Physique……...…72

3.3 Domaine de la Psychologie……...……74
3.3.1 Test de Stroop……74
3.3.2 Banque d'Images…….....…76

3.4 Synthèse des Travaux Expérimentaux……..79

3.5 Paramètres Acoustiques Mesurés……81
3.5.1 Paramètres Classiques……81
3.5.2 La Fréquence d'Équilibre Spectral……87
3.5.3 Le « Digital Amplitude Length »……95
3.5.4 La Modulation d'Amplitude……101
3.5.5 La Modélisation……106

3.6 Synthèse des Travaux Théoriques……112

4. Conclusion……115

5. Références Bibliographiques……121

5.1 Publications de l'Auteur……121

5.2 Publications de la Revue Générale des Connaissances……124

5.2.1 Analyse de la Voix dans le Secteur des Transports……124
5.2.2 Voix et Émotion……126
5.2.3 Voix et Fatigue……139
5.2.4 Voix et Autres Facteurs Perturbateurs……141
5.2.5 Ouvrages Généraux et Publications Complémentaires……142

6. Table des Matières……144