Sujet de stage M2

Etude de paramètres de qualité de voix pour la détection de la dépression

Encadrement :
- J.-L. Rouas (CR CNRS, LaBRI)
- P. Philip (PU-PH, Sanpsy)

Sujet résumé :
L'analyse de la voix, et particulièrement des aspects prosodiques (rythme, intonation, hauteur, qualité vocale), est un moyen de caractériser les états émotionnels des individus. De nombreux articles scientifiques récents tendent à montrer que les états émotionnels des locuteurs peuvent être déterminés par le biais de l'analyse de la voix (notamment en ce qui concerne les états de somnolence et de dépression). Cependant, malgré l'assurance que nous apportent ces publications sur la pertinence de la modélisation de la voix pour la caractérisation de l'humeur ou du niveau d'éveil, un certain nombre de problèmes sont toujours présents : - Quelle base de données utiliser pour la validation (comment contrôler l'acquisition des données et les valider par des experts) ? - Comment définir un ensemble restreint de paramètres pertinents ?

L'objectif de ce stage est d'apporter des éléments de réponse à ces questions. Nous allons procéder à des enregistrements contrôlés cliniquement de patients du CHU avec l'aide de médecins afin d'établir un diagnostic précis, avec un objectif d'une dizaine d'enregistrements par semaine au cours de la durée du stage. Il faudra mettre au point d'un protocole d'enregistrement et le faire valider par les médecins. Ce protocole devra comporter une lecture de texte et une phase d'entretien guidé. Les enregistrements auront lieu au CHU aux heures de disponibilité des patients (probablement fin de journée / soirée) et seront entièrement supervisés par le stagiaire. La constitution de cette base de données est une part importante du stage car elle permettra d'effectuer de nombreuses études postérieures.

En parallèle, nous étudierons la pertinence de plusieurs types de descripteurs pour la détection de la dépression en lien étroit avec les indices utilisés par les médecins. L'objectif est de confronter les paramètres utilisés au LABRI pour la caractérisation de la voix (paramètres cepstraux, prosodie, qualité de voix) aux paramètres utilisés dans les systèmes état de l'art (notamment les systèmes proposés dans le cadre du challenge AVEC 2016) et de déterminer un ensemble de paramètres pertinents pouvant inclure des descripteurs originaux.

Sujet détaillé :

L'analyse de la voix, et particulièrement des aspects prosodiques (rythme, intonation, hauteur, qualité vocale), est un moyen de caractériser les états émotionnels des individus. De nombreux articles scientifiques récents tendent à montrer que les états émotionnels des locuteurs peuvent être déterminés par le biais de l'analyse de la voix (notamment en ce qui concerne les états de somnolence et de dépression) [1, 2, 3, 4], avec en particulier l'établissement de challenges internationaux depuis quelques années (Challenges Interspeech depuis 2009 et AV+EC depuis 2010) et la mise à disposition de bases de données permettant d'évaluer correctement ce type de tâches. Les derniers articles parus sur le sujet utilisent les bases de données sus-citées. Par exemple [5] et [6] décrivent et analysent la pertinence de différents paramètres extraits du signal audio dans l'objectif d'identifier la parole dite "dépressive" et "somnolente", tandis que [7] et [8] traitent de la modélisation statistique permettant de créer les modèles de classification adaptés.

Cependant, malgré l'assurance que nous apportent ces publications sur la pertinence de la modélisation de la voix pour la caractérisation de l'humeur ou du niveau d'éveil, un certain nombre de problèmes sont toujours présents :
- Quelle base de données utiliser pour la validation (comment contrôler l'acquisition des données et les valider par des experts) ?
- Comment définir un ensemble restreint de paramètres pertinents ?
- Quels modèles utiliser pour la classification automatique ?

Les solutions généralement apportées par la communauté sont les suivantes :
- La mise en place de challenges internationaux permettant la validation de systèmes de caractérisation automatique sur des jeux de données communs. Nous pouvons ici citer par exemple le challenge Interspeech « Speaker State Challenge » de 2011 [7] pour lequel deux bases de données ont été utilisées : les corpus ALC « Alcohol Language Corpus » (parole « alcoolisée ») et SLC « Sleepy Language Corpus » (parole « somnolente ») contenant un nombre conséquent de locuteurs allemands (respectivement 162 et 99) et le challenge AV+EC 2015 [5] qui utilise la base de données RECOLA contenant des enregistrements audio, vidéo et de capteurs physiologiques de 27 locuteurs [6]. Les principaux désavantages liés à ces données sont : les domaines couverts sont déterminés par les organisateurs du challenge, le manque de contrôle sur les conditions (notamment cliniques) d'enregistrement des données.
- L'extraction du signal audio de paramètres de plus en plus nombreux permettant d'avoir la couverture la plus large possible (i.e. Challenge Interspeech 2009 Emotion Challenge : 384 paramètres [7], Interspeech 2010 Paralinguistic Challenge : 1582 paramètres [8], Interspeech 2011 Sleepiness Sub-Challenge : 4368 paramètres [9]). Cette solution n'est pas entièrement satisfaisante car malgré son efficacité se pose les questions liées au temps de calcul nécessaire à l'extraction et à l'identification de paramètres pertinents pour une tâche spécifique.
- Les modèles généralement utilisés pour la classification automatique sont des modèles dont l'efficacité est reconnue comme les modèles de mélange de lois gaussiennes (GMM) ou les machines à vecteurs supports (SVM) [10, 11]. Très peu de recherches sont menées sur cette problématique dans le cadre des états affectifs.

L'objectif de ce stage est d'apporter des éléments de réponse aux deux premières questions. Nous allons procéder à des enregistrements contrôlés cliniquement de patients du CHU avec l'aide de médecins afin d'établir un diagnostic précis, avec un objectif d'une dizaine d'enregistrements par semaine au cours de la durée du stage. Il faudra mettre au point d'un protocole d'enregistrement et le faire valider par les médecins. Ce protocole devra comporter une lecture de texte et une phase d'entretien guidé. Les enregistrements auront lieu au CHU aux heures de disponibilité des patients (probablement fin de journée / soirée) et seront entièrement supervisés par le stagiaire. La constitution de cette base de données est une part importante du stage car elle permettra d'effectuer de nombreuses études postérieures.

En parallèle, nous étudierons la pertinence de plusieurs types de descripteurs pour la détection de la dépression en lien étroit avec les indices utilisés par les médecins. L'objectif est de confronter les paramètres utilisés au LABRI pour la caractérisation de la voix (paramètres cepstraux, prosodie, qualité de voix) aux paramètres utilisés dans les systèmes état de l'art (notamment les systèmes proposés dans le cadre du challenge AVEC 2016 [12-16]) et de déterminer un ensemble de paramètres pertinents pouvant inclure des descripteurs originaux.

Références :

[1] Hönig, F.; Batliner, A.; Nöth, E.; Schnieder, S.; Krajewski, J. "Automatic Modelling of Depressed Speech: Relevant Features and Relevance of Gender". In: Proc of Interspeech 2014, Singapore, 14-18 September 2014.

[2] Cummins, N.; Sethu, V.; Epps, J.; Schnieder, S.; Krajewski, J. "Analysis of acoustic space variability in speech affected by depression". In: Speech Communication, vol. 75, pp 27-49, 2015.

[3] Sturim, D.; Torres-Carrasquillo, P.; Quatieri, T.F.; Malyska, N.; McCree, A. "Automatic Detection of Depression in Speech using Gaussian Mixture Modeling with Factor Analysis". In: Proc. of Interspeech 2011, Florence, Italy, 27-31 August 2011.

[4] Cummins, N.; Sethu, V.; Epps, J.; Krajewski, J. "Relevance Vector Machine for Depression Prediction". In: Proc. of Interspeech 2015, Dresden, Germany, 6-10 September 2015.

[5] Fabien Ringeval, Björn Schuller, Michel Valstar, Shashank Jaiswal, Erik Marchi, Denis Lalanne, Roddy Cowie, and Maja Pantic. 2015. AV+EC 2015: The First Affect Recognition Challenge Bridging Across Audio, Video, and Physiological Data. In Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge (AVEC '15). ACM, New York, NY, USA, 3-8.

[6] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In Proc. Of EmoSPACE, FG , Shanghai, China, 2013.

[7] Schuller, B., Steidl, S., Batliner, A., 2009. The Interspeech 2009 emotion challenge. In: Tenth Annual Conference of the International Speech Communication Association.)

[8] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S., 2010. The INTERSPEECH 2010 paralinguistic challenge.In: Eleventh Annual Conference of the International Speech Communication Association.

[9] Schuller, B., Steidl, S., Batliner, A., Schiel, F., Krajewski, J., 2011. The INTERSPEECH 2011 Speaker State Challenge. In: Interspeech (2011).ISCA, Florence, Italy.

[10] Dong-Yan Huang, Zhengchen Zhang, Shuzhi Sam Ge, Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines, Computer Speech & Language, Volume 28, Issue 2, March 2014, Pages 392-419,

[11] James R. Williamson, Thomas F. Quatieri, Brian S. Helfer, Gregory Ciccarelli, and Daryush D. Mehta. 2014. Vocal and Facial Biomarkers of Depression based on Motor Incoordination and Timing. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC '14). ACM, New York, NY, USA, 65-72.

[12] Williamson, J. R.; Godoy, E.; Cha, M.; Schwarzentruber, A.; Khorrami, P.; Gwon, Y.; Kung, H.-T.; Dagli, C. & Quatieri, T. F. Detecting Depression Using Vocal, Facial and Semantic Communication Cues Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 2016, 11-18

[13] Huang, Z.; Stasak, B.; Dang, T.; Wataraka Gamage, K.; Le, P.; Sethu, V. & Epps, J. Staircase Regression in OA RVM, Data Selection and Gender Dependency in AVEC 2016 Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 2016, 19-26

[14] Pampouchidou, A.; Simantiraki, O.; Fazlollahi, A.; Pediaditis, M.; Manousos, D.; Roniotis, A.; Giannakakis, G.; Meriaudeau, F.; Simos, P.; Marias, K.; Yang, F. & Tsiknakis, M. Depression Assessment by Fusing High and Low Level Features from Audio, Video, and Text Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 2016, 27-34

[15] Ma, X.; Yang, H.; Chen, Q.; Huang, D. & Wang, Y. DepAudioNet: An Efficient Deep Model for Audio Based Depression Classification Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 2016, 35-42

[16] Nasir, M.; Jati, A.; Shivakumar, P. G.; Nallan Chakravarthula, S. & Georgiou, P. Multimodal and Multiresolution Depression Detection from Speech and Facial Landmark Features Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, ACM, 2016, 43-50