Vers un système de production de feedbacks personnalisé pour l'entraînement de la prise de parole en public

Nesrine Fourati, Mathieu Chollet

▶ **To cite this version:**

Nesrine Fourati, Mathieu Chollet. Vers un système de production de feedbacks personnalisé pour l'entraînement de la prise de parole en public. Workshop sur les Affects, Compagnons artificiels et Interactions, Jun 2020, Saint Pierre d'Oléron, France. hal-02933491

HAL Id: hal-02933491
https://inria.hal.science/hal-02933491

Submitted on 8 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vers un système de production de feedbacks personnalisé pour l'entraînement de la prise de parole en public

Nesrine Fourati
nfourati@umanis.com
Umanis Research & Innovation
Levallois-Perret, France

Mathieu Chollet
mathieu.chollet@imt-atlantique.fr
IMT Atlantique
Nantes, France

RÉSUMÉ

Motivés par le besoin de former les collaborateurs d'entreprises du secteur industriel, nous présentons un système de production de feedbacks pour l'entraînement de la prise de parole en public. Dans un premier temps, la description de la performance est réduite à l'expression vocale.

KEYWORDS

Prise de Parole en Public, Entraînement Social

ACM Reference Format:
Nesrine Fourati and Mathieu Chollet. 2020. Vers un système de production de feedbacks personnalisé pour l'entraînement de la prise de parole en public. In *WACAI 2020: Affects, Compagnons Artificiels et Interaction, 03-05 Juin, 2020, Ile d'Oléron.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Les compétences sociales sont cruciales dans un grand nombre de professions, et sont régulièrement citées comme certaines des compétences les plus prisées par les recruteurs. La Prise de Parole en Public (PPP), est particulièrement importante de par sa prévalence. Une des méthodes classiques pour l'amélioration de ses compétences en PPP consiste à suivre les recommandations illustrées dans des livres spécialisés décrivant les éléments permettant de réaliser une bonne performance de PPP. Cependant, il est difficile pour un apprenant de juger sa propre performance. Il n'est donc pas surprenant que de nombreuses formations en présentiel existent pour s'entraîner à la PPP. Si la présence d'un formateur est privilégiée pour ce type de scénario, il est malheureusement impossible d'en avoir recours de façon permanente ; ce problème est en particulier fréquemment évoqué par les collaborateurs dans le secteur industriel ayant pu suivre une session de formation en PPP de courte durée (souvent de deux-trois jours) sans suivi ultérieur. Motivés par les limites de ces approches "classiques" et par la maturité des système d'acquisition et de modélisation automatique des signaux sociaux, plusieurs travaux de recherche et outils commerciaux se sont intéressés à l'entraînement de la prise de parole en public [1, 3, 10]. Cependant, de nombreuses questions de recherche restent ouvertes sur la mise en place de l'approche pédagogique dans de tels outils et des paramètres permettant de les rendre les plus efficaces. Dans cet article, nous présentons nos travaux préliminaires sur la constitution d'un outil d'entraînement à la prise de parole en public et nous évoquons les questions de recherche auxquelles nous nous attaquons.

2 ÉTAT DE L'ART

De nombreux outils de production de feedbacks pour l'entraînement de la Prise de Parole en Public (PPP) commencent à être développés. Différentes techniques pédagogiques sont régulièrement adoptées pendant les formations animées par les experts. Ces approches sont souvent adaptées pour la création des systèmes automatiques pour l'entraînement de la PPP. Parmi les approches pédagogiques les plus utilisées, nous pouvons citer le *modeling*, c'est à dire la démonstration d'exemples à suivre ou à ne pas suivre, et le *feedback*, c'est à dire la production d'évaluations et de retours personnalisés.

Certains outils récemment développés intègrent à la fois le *modeling* et le *feedback* [10], tandis que d'autres se focalisent uniquement sur une de ces deux approches [9]. Une nouvelle vague de travaux s'intéresse à des simulations de situations de PPP, mettant en jeu un public de personnages virtuels [4]. L'objectif principal de la mise en place d'une audience virtuelle est la réduction du stress et l'anxiété engendrée par une PPP. D'autres études se sont focalisées sur la production de *feedbacks* détaillés constituant une sorte de transcription du comportement social non-verbal du participant, souvent inconsciemment exprimé [5, 9, 10]. L'objectif principal d'une telle approche consiste à aider le participant à monter en compétences en PPP en lui fournissant un moyen de s'auto-évaluer. Nous avons choisi de nous focaliser dans un premier temps sur cette approche. L'analyse automatique d'un comportement social peut se reposer sur un ou plusieurs signaux sociaux comme l'expression vocale, faciale, ou corporelle. Plusieurs chercheurs ont essayé de caractériser automatiquement la qualité d'un discours pendant une PPP [3, 7, 8, 10]. Certaines études récentes commencent à s'intéresser à l'expressivité corporelle pendant une PPP [1, 9]. D'autres travaux récents se sont basées sur une analyse multimodale [11]. Nous nous focalisons dans la première version

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WACAI 2020, 03–05 Juin, 2020, Ile d'Oléron
© 2020 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . .$15.00
https://doi.org/10.1145/1122445.1122456

de notre outil sur le signal vocal présentant la modalité de communication principale pendant une PPP. Les questions de recherche que nous souhaitons aborder sont principalement : 1) Quelles sont les caractéristiques du signal vocal, "interprétables" par l'utilisateur, permettant de décrire une bonne performance de PPP, 2) Comment présenter ces caractéristiques par une visualisation adaptée pour sensibiliser le sujet sur sa performance vocale ?, et 3) A quel point notre outil pourrait être bénéfique pour le participant pour monter en compétences en PPP ?

3 SYSTÈME

Dans cette section, nous présentons notre système de production de feedbacks personnalisé pour l'entraînement de la PPP. Dans le contexte de ce projet, nous avons suivi un groupe de personnes travaillant dans le milieu industriel, et ayant été sollicitées pour suivre une formation en PPP étalée sur deux jours et animée par un coach professionnel (directrice développement RH spécialisée dans les compétences sociales requises pour la réussite d'une PPP).

Analyse acoustique : Nous avons fait le choix de favoriser des paramètres acoustiques prédéfinis à l'avance et interprétables par les participants. Après avoir échangé avec le coach et en nous inspirant des études effectuées récemment [10], nous proposons dans un premier temps de se focaliser sur trois caractéristiques vocales : les pauses, le débit de parole et la variation du Pitch. Contrairement à d'autres travaux qui se basent sur des outils commerciaux pour caractériser un signal audio pendant une PPP [10], nous proposons d'utiliser une nouvelle librairie récemment présente en Python permettant d'exécuter les fonctionnalités de Praat telles qu'elles sont [6]. Nous optons pour cette solution afin de profiter des avantages proposés par le logiciel Praat, une référence pour l'analyse et la transcription phonétique [2].

Le type de Feedback : Nous proposons de nous focaliser sur trois aspects de feedback. Le premier aspect porte sur une transcription automatique du comportement non-verbal à travers la visualisation des trois caractéristiques de l'expression vocale que nous avons sélectionnées. Le deuxième aspect porte sur l'évolution de la performance du sujet tout au long de la formation. Ceci lui permet de se rendre compte de sa capacité à intégrer les consignes apprises avec des approches pédagogiques données par le coach. Le troisième aspect couvre le positionnement de sa performance par rapport aux autres participants ayant assisté à la même formation. Par exemple, le système pourrait afficher la moyenne du groupe ainsi que d'autres valeurs de référence définies par le coach. Cela permet d'ajouter une couche informative d'interprétation du feedback fourni au participant.

Visualisation : La visualisation des données est considérée comme une approche intuitive et efficace pour transmettre facilement et rapidement toutes sortes d'information, aussi complexes qu'elles pourraient être, y inclus celles représentant les signaux sociaux [1, 10, 11]. Pour une observation donnée, nous affichons les trois caractéristiques de l'expression vocale, à savoir la variation du Pitch, les pauses et le débit de parole, sous deux formes de représentation : globale et locale. La représentation globale permet de résumer les valeurs décrivant ces caractéristiques tout au long du discours (e.g. le pourcentage de la durée totale de silences présents dans le discours). La représentation locale permet d'avoir une meilleure visibilité sur l'évolution temporelle de ces caractéristiques (e.g. le positionnement et la durée des différentes pauses).

4 CONCLUSION

Pour la première version de notre outil d'entraînement à la PPP, nous nous focalisons sur quelques caractéristiques interprétables décrivant l'expression vocale. Même si la visualisation vulgarisée de ces caractéristiques semble être amplement appréciée d'après le retour de quelques collaborateurs, nous comptons mettre en place dans le futur une expérimentation afin de valider et raffiner notre plate-forme auprès d'un groupe de participants tiers. Nous comptons aussi enrichir notre système pour rendre l'interface interactive, pour considérer d'autres approches pédagogiques comme le *modeling*, pour inclure d'autres aspects décrivant l'expressivité vocale, ainsi que d'autres signaux sociaux comme la posture et les gestes. Une approche itérative mettant en place l'interaction entre le retour d'expérience de l'utilisateur et la mise à jour de l'outil est envisagée dans nos futurs travaux.

RÉFÉRENCES

[1] Roghayeh Barmaki. 2016. Improving Social Communication Skills Using Kinesics Feedback. In *Proc. of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM Press, New York, New York, USA, 86–91.

[2] Paul Boersma and David Weenink. 2018. Praat : doing phonetics by computer [Computer program] Version 6.0.37, retrieved 14 March 2018 from http ://www.praat.org. (2018).

[3] Mathieu Chollet and Stefan Scherer. 2017. Assessing Public Speaking Ability From Thin Slices of Behavior. In *Proc. of the IEEE Int. Conference on Face and Gesture Recognition*.

[4] Mathieu Chollet, Torsten Wortwein, Louis-Philippe Morency, Ari Shapiro, and Stefan Scherer. 2015. Exploring Feedback Strategies to Improve Public Speaking : An Interactive Virtual Audience Framework. In *Proc. of UbiComp 2015*. ACM, Osaka, Japan.

[5] Michelle Fung, Yina Jin, RuJie Zhao, and Mohammed Hoque. 2015. ROC speak : semi-automated personalized feedback on nonverbal behavior from recorded videos. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1167–1178.

[6] Yannick Jadoul, Bill Thompson, and Bart de Boer. 2018. Introducing Parselmouth : A Python interface to Praat. *Journal of Phonetics* 71, 2018 (2018), 1–15.

[7] A. Rosenberg and J. Hirschberg. 2005. Acoustic/prosodic and lexical correlates of charismatic speech. In *Proceedings of Interspeech 2005*. 513–516.

[8] E. Strangert and J. Gustafson. 2008. What Makes a Good Speaker ? Subject Ratings, Acoustic Measurements and Perceptual Evaluations. In *Proceedings of Interspeech 2008*. 1688–1691.

[9] M Iftekhar Tanveer, Ru Zhao, Kezhen Chen, Zoe Tiet, and Mohammed Ehsan Hoque. 2016. AutoManner : An Automated Interface for Making Public Speakers Aware of Their Mannerisms. In *Proc. of the 2016 ACM Conference on Intelligent User Interfaces*.

[10] Xingbo Wang, Haipeng Zeng, Yong Wang, Aoyu Wu, Zhida Sun, Xiaojuan Ma, and Huamin Qu. 2020. VoiceCoach : Interactive Evidence-based Training for Voice Modulation Skills in Public Speaking. (2020). arXiv:2001.07876

[11] Haipeng Zeng, Xingbo Wang, Aoyu Wu, Yong Wang, Quan Li, Alex Endert, and Huamin Qu. 2019. EmoCo : Visual Analysis of Emotion Coherence in Presentation Videos. *IEEE Transactions on Visualization and Computer Graphics* (2019), 1–1.