# Enseignement de l'intonation du français par une synthèse vocale contrôlée par le geste : étude de faisabilité

Xiao Xiao, Corinne Bonnet, Haohan Zhang, Nicolas Audibert, Barbara Kühnert, Claire Pillot-Loiseau

# Enseignement de l'intonation du français par une synthèse vocale contrôlée par le geste : étude de faisabilité

Xiao Xiao[1] Corinne Bonnet[2] Haohan Zhang[3] Nicolas Audibert[3] Barbara Kühnert[3] Claire Pillot-Loiseau[3]

(1) Léonard de Vinci Pôle Universitaire, Research Center, 12, avenue Léonard de Vinci, Paris – La Défense 92400, France

(2) DILTEC Didactique des langues, des textes et des cultures EA 2288, Sorbonne Nouvelle, 4, rue des Irlandais, 75005 Paris, France, et Université Toulouse Paul Sabatier, 118 Rte de Narbonne, 31062 Toulouse, France

(3) Laboratoire de Phonétique et Phonologie UMR 7018, Sorbonne Nouvelle, 4, rue des Irlandais, 75005 Paris, France

xiao.xiao@devinci.fr, {corinne.bonnet, haohan.zhang nicolas.audibert, barbara.kuhnert, claire.pillot}@sorbonne-nouvelle.fr

## RESUME

Peut-on enseigner l'intonation française en classe avec une synthèse vocale contrôlée gestuellement sur une tablette ? La fréquence fondamentale et la durée de quatre phrases déclaratives, quatre questions polaires, quatre énoncés exprimant l'incrédulité (1 à 4 syllabes) de deux apprenantes ukrainiennes débutantes en français ont été comparées avant et après quatre entraînements hebdomadaires. Les apprenantes devaient écouter un enregistrement de référence, puis visualiser le modèle sur la tablette, tracer l'intonation manuellement, écouter le résultat synthétisé, et tracer et écouter leur tracé sans guide. Elles produisaient initialement des phrases déclaratives avec une intonation ascendante, et ont différencié les déclarations et les questions polaires après l'entraînement. L'expression de l'incrédulité s'est améliorée pour l'une. L'autre a montré quelques difficultés à maîtriser cette technologie. Cette première étude de cas utilisant la synthèse vocale contrôlée gestuellement est une approche prometteuse permettant plus de pratique de l'intonation en classe.

## ABSTRACT

**Teaching French intonation using gesture-controlled speech synthesis: a feasibility study.**

Is learning French intonation with gesture-controlled speech synthesis on a tablet in the classroom feasible? The fundamental frequency and duration of four declarative sentences, four polar questions, four statements expressing incredulity (1 to 4 syllables) of two Ukrainian beginner learners of French were compared before and after four weekly training sessions. Learners were asked to listen to a reference recording, then look at the intonation guide on the tablet, trace the intonation manually, listen to the synthesized result, and trace and listen to their tracing without a guide. They initially produced declarative sentences with rising intonation and differentiated statements and polar questions after training. The expression of disbelief improved for one of them. The other showed some difficulty in mastering this technology. This first case study using gesture-controlled speech synthesis is a promising approach for more intonation practice in the classroom.

MOTS-CLES : synthèse vocale contrôlée par le geste, intonation, français, salle de classe

KEYWORDS : gesture-controlled vocal synthesis, intonation, French, classroom setting

# 1 Introduction

L'acquisition de l'intonation peut s'avérer difficile pour les apprenants non natifs (Mennen, 2015). Plusieurs stratégies d'enseignement ont été proposées et testées, visant à attirer l'attention sur les changements de hauteur pour la perception et la production. L'association entre un son et une représentation visuelle ou gestuelle de son contour de fréquence fondamentale ($f_o$) peut aider un apprenant à percevoir un mouvement $f_o$ non familier et à ancrer l'établissement de nouvelles catégories de mouvements de hauteur (Yuan et al., 2019). Les représentations visuelles des contours $f_o$ peuvent également transmettre les différences entre un modèle et les productions des apprenants, pour les aider à comprendre et à corriger leurs propres erreurs (Taniguchi et Abberton, 1999). Pour la production, les gestes ont été utilisés pour renforcer kinesthésiquement les caractéristiques de l'intonation (entre autres : Bailly et al., 2022).

Notre recherche explore comment la synthèse vocale contrôlée en temps réel par des gestes de la main, appelée Synthèse Vocale Performative (PVS, Locqueville et al., 2020 ; Xiao et al., 2023), peut être utilisée par des locuteurs non-natifs pour la pratique de l'intonation d'une langue étrangère. Des études pilotes de PVS avec des apprenants de L2 ont été menées sur des corpus français et anglais, à l'aide de l'interface Gepeto sur tablette mobile. Ces premières études ont permis de valider le potentiel de l'utilisation de PVS pour l'apprentissage de l'intonation (Xiao et al., 2021, 2023). Auparavant, le déploiement de Gepeto était limité à des études à session unique dans des conditions de laboratoire contrôlées. Le présent travail explore la manière dont Gepeto peut être incorporé de manière longitudinale pour compléter l'enseignement des langues étrangères en classe.

Dans cet article, nous présenterons une étude de cas sur l'acquisition du français par des apprenants ukrainiens dont l'une des difficultés de communication réside dans la compréhension et la production de la distinction prosodique entre les énoncés et les questions polaires (questions oui/non) en français. Certaines de ces difficultés pourraient être dues aux différences entre les systèmes prosodiques du français et de l'ukrainien : l'ukrainien est une langue slave avec une accentuation lexicale libre. Les phrases déclaratives et les questions "wh" sont caractérisées par un contour d'intonation descendant, tandis que les questions polaires sont produites avec un contour d'intonation ascendant. Cependant, l'ukrainien possède également des accents de hauteur qui signalent une focalisation large : dans ce cas, ils sont produits avec un accent prénucléaire ascendant (bas-haut) suivi d'un accent nucléaire descendant (haut-bas). Pour ces raisons, certaines phrases déclaratives peuvent présenter des schémas intonatifs ascendants (Pompino-Marschall et al., 2017). Le français est une langue romane avec un accent sur la dernière syllabe d'un groupe de mots (Fougeron et Smith, 1993). En français, sans expression d'une attitude particulière, les phrases déclaratives se terminent par un contour descendant, tandis que les questions polaires se terminent avec un contour ascendant (Di Cristo, 2016).

Cette étude de cas cherche à savoir si l'entraînement à l'utilisation de la PVS permet aux apprenants ukrainiens débutants de français d'améliorer leur production différenciée de phrases déclaratives, de questions polaires et d'un modèle d'intonation attitudinal (incrédulité) dans des énoncés courts.

# 2 Matériel et Méthodes

L'interface Gepeto a été utilisée pour les conditions gestuelles de l'étude. Elle consiste en une interface mobile personnalisée qui contrôle un synthétiseur vocal, Voks, qui permet le contrôle mélodique et rythmique en temps réel d'échantillons préalablement enregistrés ou de synthèse vocale par l'utilisation de gestes de la main (Locqueville et al., 2020). L'interface mobile fonctionne dans le navigateur d'une tablette ou d'un téléphone portable. Cette étude a utilisé les téléphones mobiles

personnels de chaque sujet, contrôlés par des mouvements du bout des doigts. Le tracé d'une courbe dans la zone de contrôle de l'interface du téléphone portable produit une resynthèse en temps réel de la phrase actuelle à partir de Voks. L'axe horizontal détermine la position temporelle de l'échantillon original à resynthétiser, et le moment de la resynthèse est déterminé par le moment du geste de l'utilisateur. L'axe vertical de la région de contrôle module la hauteur de la sortie. Il est régulièrement espacé sur une échelle de demi-tons (ST) avec une gamme de 24ST (2 octaves) calibrée autour du corpus de l'étude. Un panneau de boutons apparaît à gauche de la région de contrôle. Le bouton du haut permet de basculer entre le mode fondu et le mode maintenu pour la trace de l'utilisateur, où la trace disparaît après 1,5 seconde ou reste jusqu'à ce qu'elle soit effacée. En mode maintenu, trois autres boutons sont activés, permettant à l'utilisateur de lire, d'effacer ou d'enregistrer le tracé du geste en cours. Le bouton situé dans le coin inférieur gauche déclenche la lecture de l'enregistrement de référence pour la phrase en cours. L'enseignant dispose d'une interface de commande distincte qui lui permet de sélectionner la phrase en cours et d'activer ou de désactiver le guide visuel montrant la courbe d'intonation de la phrase de référence.

Le profil linguistique des apprenants a été documenté via un questionnaire initial : langue maternelle et autres langues apprises, durée du séjour en France et pratique du français, niveau de français, formation dans leur pays et en France, et utilisation actuelle du français et de la langue maternelle.

## 2.1 Contexte général de l'étude

Une classe d'étudiants ukrainiens adultes de niveau débutant a été sélectionnée pour l'étude. Les sujets ont été recrutés dans le cadre d'un cours d'introduction au français qui s'est déroulé au printemps 2023 à l'Université Paul Sabatier de Toulouse. Le cours d'1,5 heure a eu lieu pendant 14 semaines consécutives et a couvert la grammaire française de base et le vocabulaire lié à la vie quotidienne. La classe suivait un programme standard, dont la partie prononciation ne couvrait que la phonétique segmentale, mais pas l'intonation. Notre intervention s'est déroulée sur 6 séances hebdomadaires, dont 4 séances d'entraînement (20-30 minutes) à l'utilisation de Gepeto. Des enregistrements d'évaluation ont été effectués lors de la première et de la dernière séance. Les sessions de formation ont été supervisées par une formatrice en langues étrangères, qui a guidé les étudiants pendant les différents exercices, les a aidés avec les aspects techniques de l'interface et a pris des notes détaillées sur ses observations et ses échanges avec les apprenants. L'entraînement et les enregistrements d'évaluation ont utilisé un corpus d'énoncés courts (1-4 syllabes) produits avec des intonations différentes selon le contexte de communication (affirmation, question ou incrédulité).

Les interventions se sont déroulées sur 6 sessions de cours dans une salle séparée à côté de la salle de classe. La première et la dernière séance ont été consacrées aux pré-tests et aux post-tests, et les séances intermédiaires ont consisté en des activités destinées à entraîner les schémas d'intonation. Les sujets se rendaient individuellement dans la salle avant ou après le cours et réalisaient les activités prévues avec l'expérimentateur. Les enregistrements d'évaluation avaient pour but d'identifier les points de difficulté et de mesurer les améliorations apportées par l'entraînement.

## 2.2 Sujets

Les étudiants étaient des ressortissants ukrainiens installés en France entre 1,5 et 12 mois avant l'étude. 7 femmes et 3 hommes ont participé au pré-test (19-65 ans, moyenne 29,5 ans). L'ukrainien était leur première langue et l'anglais ou le russe (pour 7 sujets) leur deuxième ou troisième langue. Tous les sujets étaient débutants en français, avec un niveau entre A0 et A2 (Conseil de l'Europe, CECR 2001). Seules deux personnes ont participé à toutes les sessions d'entraînement : UF1 (femme,

37 ans, vivant en France depuis 1 an, niveau de français A1, chercheuse) et UF2 (femme, 65 ans, vivant en France depuis 6 mois, niveau de français A0, chimiste). Leurs deuxième et troisième langues apprises sont respectivement l'anglais et le russe. Les deux apprenantes n'ont aucune pratique musicale et n'ont jamais eu d'entraînement préalable à la mélodie du français parlé.

## 2.3 Corpus

Le corpus se compose de mots et de phrases françaises courtes de 1 à 4 syllabes enregistrées par un locuteur natif masculin avec trois modalités distinctes associées à des schémas d'intonation différents : affirmation (déclaration), question ou incrédulité. Le corpus a été divisé en trois groupes. Les énoncés du groupe A ont été utilisés pendant le pré-test ; ceux du groupe B lors des sessions de formation, et ceux du groupe C lors du post-test. Nous avons choisi des énoncés courts pour éviter les problèmes liés au rythme des phrases plus longues que nous avons rencontrés dans nos études précédentes (Xiao et al., 2021). De plus, les énoncés courts sont plus faciles à reproduire pour les apprenants de français de niveau débutant. Enfin, différents mots ont été enregistrés dans le pré- et le post-test afin de tester la capacité des apprenants à généraliser et à appliquer les modèles d'intonation qu'ils ont appris à d'autres exemples. Les enregistrements ont été réalisés sur un ordinateur portable à l'aide d'un microphone de bureau unidirectionnel à condensateur en utilisant le logiciel Audacity.

## 2.4 Analyse acoustique

La fréquence fondamentale ($f_o$) a été extraite à l'aide de scripts Praat (Boersma et Weenink 2023). Les valeurs minimales et maximales possibles ont été fixées en fonction de la portée observée du locuteur afin de minimiser les erreurs de détection. Les contours $f_o$ obtenus ont été inspectés pour éliminer les erreurs de détection. Après conversion en demi-tons, la position temporelle des points de mesure et les valeurs $f_o$ associées ont été extraites.

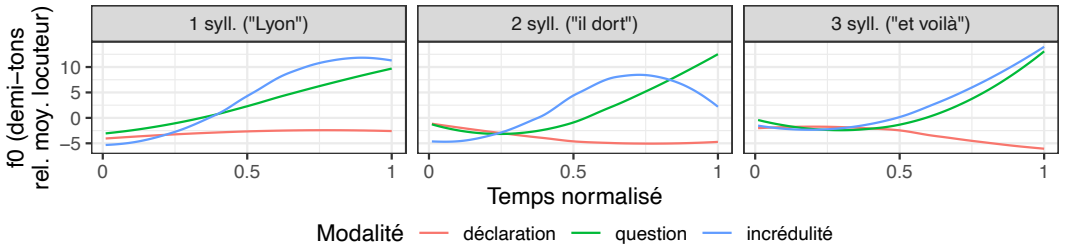


FIGURE 1 : Courbes de fréquence fondamentale ($f_o$) pour un énoncé monosyllabique, un disyllabique et un trisyllabique produit par le locuteur natif masculin dans chacune des trois modalités et utilisés dans les sessions d'entraînement (groupe B).

En outre, les enregistrements audio ont été segmentés en syllabes et en phones à l'aide de WebMAUS (Kisler et al., 2017). La segmentation obtenue a été corrigée manuellement, puis exportée au format TextGrid de Praat pour être affichée via l'interface Gepeto. La figure 1 montre des exemples de courbes $f_o$ pour chaque modalité du groupe B (énoncés utilisés dans les sessions d'entraînement) produite par le sujet natif, en fonction de la longueur en syllabes de l'énoncé.

L'analyse de $f_o$ de chaque énoncé produit par les apprenants lors du pré-test et du post-test a suivi la même procédure semi-automatique que celle décrite ci-dessus. La durée de chaque production a été extraite de la segmentation. Étant donné le nombre limité de sujets et d'exemples, l'analyse acoustique

se concentre principalement sur l'inspection visuelle de la forme du contour intonatif, considérée individuellement ou en moyenne sur les sujets, les conditions et les durées d'énonciation.

### 2.5 Prétest et Posttest

Des enregistrements ont été réalisés lorsque les sujets prononçaient les mots et les phrases pendant le pré-test et le post-test à l'aide d'un microphone d'ordinateur portable et du logiciel Audacity. Les instructions concernant les enregistrements étaient affichées sur des diapositives sur un écran d'ordinateur. Par exemple, pour le pré-test, les apprenants devaient produire le mot normalement comme s'il était à la fin d'une phrase ou question, en faisant attention à la ponctuation. Les modèles d'intonation des énoncés et des questions polaires étaient indiqués par des signes de ponctuation en rouge à la fin de chaque mot ou phrase. Les instructions ont été données en anglais, le niveau de français des apprenants débutants n'étant pas encore suffisant. Afin de fournir un contexte approprié pour l'intonation associée à l'incrédulité, une phrase d'introduction a été lue par l'expérimentateur pour chaque énoncé avant la production des stimuli par les sujets (exemple pour les stimuli « Lille ?! », « votre ami vous annonce qu'il déménage à Lille, à 900 km de Toulouse, et vous ne le croyez pas »).

### 2.6 Entraînement

Toutes les sessions de formation ont été supervisées par un professeur de langue et guidées par un jeu de diapositives d'instructions. Chaque session de formation utilisait un ensemble différent de 3 à 5 mots ou phrases. Les élèves ont participé aux séances de formation individuellement. Les séances d'entraînement commençaient par la lecture de chaque mot ou phrase avec les trois types d'intonation, suivie d'exercices utilisant l'interface Gepeto. Pour la tâche de lecture, les stimuli étaient présentés un par un sur une diapositive avec différents signes de ponctuation pour indiquer l'intonation appropriée (point pour les affirmations, point d'interrogation pour les questions polaires ou point d'interrogation et d'exclamation pour l'incrédulité). De plus, pour l'incrédulité, la même phrase de mise en contexte a été utilisée que dans les pré et post-tests.

Quatre types d'exercices ont été proposés avec Gepeto, sans limitation du nombre de répétitions, le guide visuel avec la production native de référence étant affiché pour les trois premiers :

1. Écouter l'enregistrement de référence tout en prêtant attention au guide visuel, puis imiter vocalement la référence.
2. Écouter la référence et la reproduire en synthèse vocale en traçant le guide visuel
3. Identique à la tâche 2 mais le geste tracé reste à l'écran et l'élève peut écouter son résultat.
4. Identique à la tâche 3, mais sans guide visuel.

Les participants écoutaient le modèle plusieurs fois avant d'essayer de reproduire les courbes. Seuls les exercices avec guides visuels ont été donnés aux élèves pour les sessions de formation 1 et 2. L'exercice sans guide a été ajouté pour les sessions de formation 3 et 4.

## 3 Résultats

### 3.1 Analyse acoustique des productions au prétest

La figure 2 montre les contours $f_o$ moyennés et normalisés dans le temps des productions des 10 apprenants ukrainiens, pour chacun des énoncés produits pendant le pré-test (groupe A du corpus) et chacune des trois modalités. Les contours $f_o$ correspondant aux productions des locuteurs natifs des

mêmes énoncés sont à droite de la figure. Les apprenants tendent à produire des contours intonatifs finaux ascendants pour les trois conditions dans le prétest, avec peu de distinction entre les modalités, bien que la variabilité entre locuteurs soit plus importante pour l'incrédulité (Figure 2). Inversement, alors que le modèle intonatif principalement produit par le locuteur natif pour exprimer l'incrédulité est proche de celui de la question sauf pour « Lille », ses énoncés déclaratifs sont systématiquement marqués par une chute intonative quel que soit le nombre de syllabes dans l'énoncé. L'incapacité apparente des apprenants ukrainiens débutants à produire la chute intonative caractéristique des énoncés déclaratifs en français peut expliquer les confusions fréquemment observées.

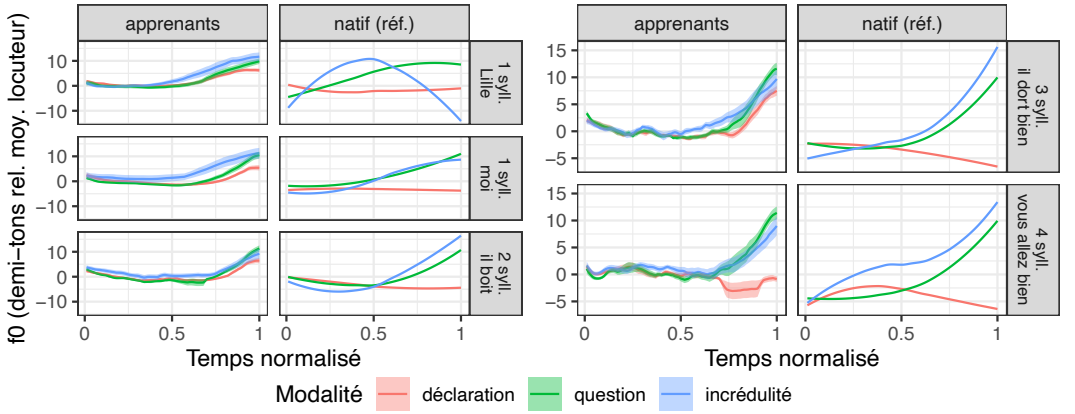


FIGURE 2 : Courbes de la production des cinq énoncés du groupe A par les 10 apprenants ukrainiens de la session de pré-test (à gauche, moyenne des apprenants), et par le locuteur natif masculin (à droite pour chaque longueur syllabique), pour chacune des trois modalités. Les enveloppes colorées autour des courbes moyennes des apprenants représentent l'erreur standard.

Une exception est l'énoncé de quatre syllabes « Vous allez bien », pour lequel les apprenants produisent en modalité déclarative des contours intonatifs plus proches de ceux du modèle natif, même s'ils présentent une trajectoire légèrement montante à la fin. Ces contours, produits avec peu de variabilité entre apprenants comme l'indique l'erreur standard modérée, est probablement influencé par les schémas prosodiques ukrainiens sur des énoncés plus longs.

## 3.2 Caractérisation acoustique des progrès des apprenantes

La figure 3 montre une comparaison entre le pré-test (groupe A) et le post-test (groupe C) des contours $f_0$ normalisés en fonction du temps dans les énoncés produits par les deux apprenantes ayant participé à l'ensemble du protocole, UF1 et UF2. Les contours $f_0$ des productions du locuteur natif (groupe A) sont reproduits à gauche. Les énoncés sont regroupés par nombre de syllabes pour la visualisation, qui comprend pour chaque locuteur et chaque condition deux énoncés monosyllabiques pour lesquels la variabilité est également représentée. Comme les énoncés de quatre syllabes n'étaient présents que dans les groupes A et B, seules les productions d'une à trois syllabes sont représentées.

La grande variabilité observée chez le locuteur natif pour les expressions d'incrédulité sur des énoncés monosyllabiques s'explique par les deux modèles intonatifs distincts produits par ce locuteur sur les énoncés « Lille » (Figure 2). Le schéma intonatif secondaire de montée et de descente se retrouve dans certaines productions d'incrédulité du post-test du locuteur UF1, qui semble avoir acquis la capacité de généraliser l'utilisation de ce schéma intonatif à des énoncés de longueur variable.

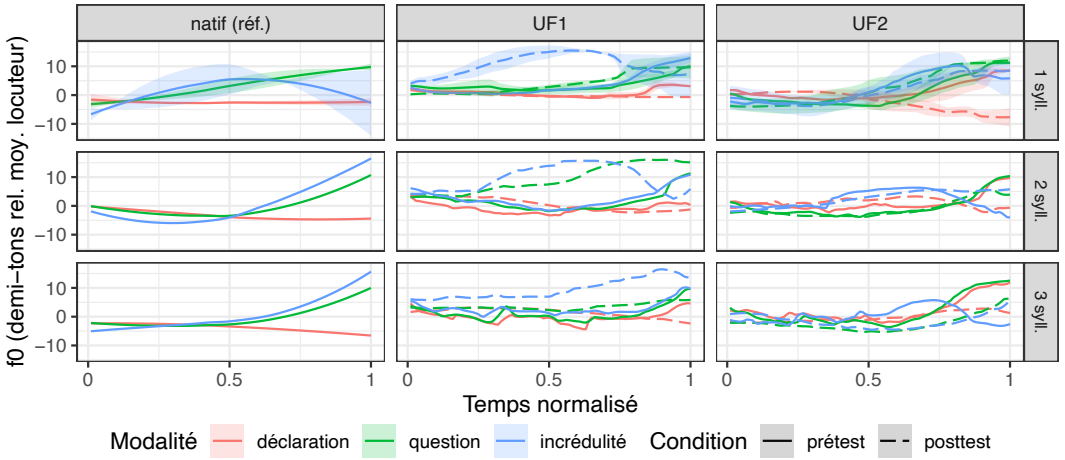


FIGURE 3 : Courbes de la production d'énoncés de 1 à 3 syllabes par les apprenants ukrainiens UF1 et UF2 qui ont participé au pré-test (énoncés du groupe A, trait plein) et au post-test (énoncés du groupe C, pointillés), pour chacune des trois modalités. Pour les énoncés monosyllabiques, des courbes moyennes sont présentées pour tenir compte des deux exemplaires inclus dans chaque condition, (enveloppes colorées : erreur standard).

En outre, les apprenants améliorent leur capacité à produire des énoncés déclaratifs et des questions de façon distincte et plus proche du natif, quel que soit le nombre de syllabes, d'où un écart plus important entre l'énoncé et la question dans la partie finale du même énoncé: l'écart moyen entre l'énoncé et la question dans le dernier quart de l'énoncé augmente entre le prétest et le post-test pour les énoncés d'une et de deux syllabes des deux apprenants (augmentation moyenne de cet écart de 9,1 demi-tons). Bien que plus modérée (3,4 demi-tons), cette augmentation de la distinction entre énoncé et question est également observée pour les énoncés de trois syllabes produits par UF1.

Enfin, la légère réduction par UF2 (-2,2 demi-tons) de l'écart entre déclaration et question (énoncés de trois syllabes), est principalement due à la baisse de $f_0$ sur la syllabe médiane lors de la production de questions dans le post-test. De plus, la forme des contours de l'énoncé et de la question produits par UF2 en fin de l'énoncé est beaucoup plus proche du natif dans le post-test que dans le pré-test.

La normalisation temporelle des courbes de $f_0$ pourrait masquer certaines distinctions entre modalités marquées par la durée et pas seulement par les variations intonatives. Si chez le natif, la distinction entre modalités n'est pas marquée par la durée, en pré-test les apprenants adoptent une stratégie de distinction par la durée (énoncés déclaratifs de 2 et 3 syllabes plus longs que les questions). En post-test, ces différences de durée sont réduites et la distinction entre les modalités s'améliore sur le plan mélodique. De plus, pour les trois modalités, quelle que soit la longueur des énoncés, les apprenants augmentent leur débit de parole entre pré- et post-test, pour les déclarations et les questions.

# 4 Discussion et conclusion

Malgré une notable réduction de la population due à des circonstances imprévues et indépendantes de notre volonté, l'étude a été utile et bénéfique pour tester la configuration expérimentale globale et la faisabilité de l'utilisation d'un système de synthèse vocale contrôlé par le geste. Elle a dévoilé plusieurs points inattendus liés à la production de l'intonation du français et à la prise en main de l'outil.

La montée finale observée pour certains des énoncés déclaratifs plurisyllabiques produits par UF2 pourrait s'expliquer par la présence éventuelle d'un accent de hauteur en ukrainien (Pompino-Marschall et al., 2017). L'intonation moyenne du pré-test pour les 10 sujets a montré un $f_0$ ascendant pour les trois types d'intonation en français, la question et l'incrédulité présentant une pente plus raide que pour la déclaration. Cependant, un schéma différent a été trouvé pour les énoncés déclaratifs de 4 syllabes, qui avaient un contour d'intonation plat dans le prétest, se terminant par une légère montée finale. Il existe donc des variations dans les modèles de production intonative en fonction de la longueur de l'énoncé. Par la suite, il serait souhaitable de tester l'outil sur des énoncés plus longs afin d'évaluer son utilité pour améliorer l'intonation de ces énoncés.

Dans le post-test, la principale amélioration s'est produite dans la production d'énoncés déclaratifs, qui ont montré un contour descendant plus fort par rapport au pré-test pour les deux apprenants. Dans l'ensemble, UF1 a réalisé des contours déclaratifs finaux plus proches de ceux du locuteur natif que UF2. UF1 semble également généraliser l'incrédulité mieux que UF2, et ses productions ont augmenté à la fin de la formation. Les différences entre les deux apprenants peuvent être en partie liées à la plus grande satisfaction d'UF1 vis-à-vis de l'expérience. De plus, UF1, plus jeune et ayant un meilleur niveau de français, a vécu plus longtemps en France que le sujet UF2. UF2 a indiqué qu'elle « parle parfois français » : on peut donc supposer que son exposition au français en dehors de la classe est limitée. Le sujet UF2 a également exprimé à plusieurs reprises sa gêne à utiliser la technologie liée à l'interface Gepeto au début des sessions de formation. Elle a été plus hésitante que UF1.

Cette étude est la première à tester une interface de contrôle gestuel dans un cadre autre que les conditions contrôlées d'un laboratoire. Parmi les difficultés rencontrées, citons les problèmes de stabilité de la connexion Internet dans la salle de classe ; la surface du téléphone en silicone, qui a entraîné une gêne pour l'un des participants qui avait les mains humides ; la latence dans l'ouverture du logiciel ; les problèmes de compatibilité entre les différents téléphones et navigateurs ; la qualité de la synthèse vocale, qui s'est parfois révélée peu familière aux participants ; et le fait que les participants n'étaient pas familiarisés avec les paramètres techniques de leur téléphone au début de l'entraînement. Dans les études futures, nous visons donc à fournir les appareils mobiles.

Malgré les différences de performance entre les deux apprenantes, toujours rencontrées en contexte de classe, et malgré le peu de sujets, cette étude exploratoire a montré qu'il est important d'enseigner l'intonation en français langue étrangère dès le début de l'apprentissage. Dans cette expérience, même s'il manque un groupe contrôle bénéficiant d'un enseignement intonatif plus « traditionnel » en comparaison avec notre pédagogie (prévu mais non réalisé en raison du blocage de l'université), l'interface Gepeto a aidé les apprenants à utiliser plus d'un modèle d'intonation en français et, surtout, à différencier les phrases déclaratives des questions polaires, une distinction qui, bien qu'essentielle à des fins de communication en français, n'est pas évidente au début pour les apprenants ukrainiens.

# Remerciements

Cette recherche a été financée par l'ANR "GEPETO" (ANR-19-CE28-0018-01), et le LabEx EFL (ANR-10-LABX-0083) qui contribue à l'IdEx Université de Paris (ANR-18-IDEX-0001). Nous remercions les enseignants et les apprenants qui ont rendu possible cette étude.

# Références


BAILLS F., ALAZARD-GUIU C. & PRIETO P. (2022). Embodied prosodic training helps improve accentedness and suprasegmental accuracy. *Applied Linguistics* 43, 776–804. https://doi.org/10.1093/applin/amac010

BOERSMA P. & WEENINK D. (2023). Praat: doing phonetics by computer. Version 6.3.06, en ligne : http://www.praat.org/

CONSEIL DE L'EUROPE. (2001). *CECRL (Cadre européen commun de référence pour les langues)*. Strasbourg : Éditions Didier.

DI CRISTO A. (2016). *Les musiques du français parlé : Essais sur l'accentuation, la métrique, le rythme, le phrasé prosodique et l'intonation du français contemporain*. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110479645

FOUGERON, C. & L. SMITH C.L. (1993). Illustrations of the IPA: French. *Journal of the International Phonetic Association 23*, 73–76.

KISLER T., REICHEL U. & SCHIEL F. (2017). Multilingual Processing of Speech Via Web Services. *Computer Speech and Language* 45, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

LOCQUEVILLE G., D'ALESSANDRO C., DELALEZ S., DOVAL B. & XIAO X. (2020). Voks: Digital Instruments for chironomic control of voice samples. *Speech Communication* 125, 97–113. https://doi.org/10.1016/j.specom.2020.10.002

MENNEN I. (2015). Beyond segments: towards a L2 intonation learning theory. In: Delais-Roussarie, E., Avanzi, M. & Herment, S. (Eds.). *Prosody and Languages in Contact: L2 acquisition, attrition, languages in multilingual situations*, Springer Verlag, pp. 171–188.

POMPINO-MARSCHALL B., STERIOPOLO E. & ŻYGIS M. (2017). Ukrainian. *Journal of the International Phonetic Association* 47, 349–57. https://doi.org/10.1017/S0025100316000372

TANIGUCHI M. & ABBERTON E. (1999). Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speach, Hearing, and Language: work in progress* 11, 77–89.

XIAO X., AUDIBERT N., LOCQUEVILLE G., D'ALESSANDRO C., KÜHNERT B., & PILLOT-LOISEAU C. (2021). Prosodic Disambiguation Using Chironomic Stylization of Intonation with Native and Non-Native Speakers. In *Proc Interspeech 2021*, pp. 516–520. https://doi.org/10.21437/Interspeech.2021-182

XIAO X., KÜHNERT B., AUDIBERT N., LOCQUEVILLE G., PILLOT-LOISEAU C., ZHANG H. & D'ALESSANDRO C. (2023). Performative Vocal Synthesis for Foreign Language Intonation Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp.1–9. https://doi.org/10.1145/3544548.3581210

YUAN C., GONZÁLEZ-FUENTE S., BAILLS F., & PRIETO P. (2019). Observing Pitch Gestures Favors the Learning of Spanish Intonation by Mandarin Speakers. *Studies in Second Language Acquisition* 41, 5–32. https://doi.org/10.1017/S0272263117000316