

# Quelle(s) mesure(s) de similarité prosodique comme évaluation de l'imitation ?

Olivier Nocaudie Corine Astésano

U.R.I. Octogone-Lordat (E.A. 4156), Université de Toulouse, UTM

nocaudie@univ-tlse2.fr, corine.astesano@univ-tlse2.fr

## RESUME

---

La performance imitative des locuteurs varie de celle du professionnel, expert, à celle du naïf, plus ou moins talentueux. L'étude de l'imitation souligne la difficulté pour trouver des indices mesurables de la réussite d'une imitation. Dans cette étude exploratoire, des contours de *f0* recueillis au fil de tâches d'imitation sont testés au moyen d'une double approche : mesure objective par le biais de deux mesures de la similarité prosodique reportées dans la littérature et évaluation perceptive par un panel de 15 auditeurs naïfs. Nos premiers résultats indiquent une bonne corrélation entre les deux approches et soulèvent la question du choix de l'indice mesurable qui rendrait le mieux compte d'une imitation au niveau tonal. Ils soulignent également la variabilité interindividuelle des comportements imitatifs en parole tout en ouvrant des perspectives intéressantes dans le domaine de la formation à la phonétique correctrice par la Méthode Verbo-tonale.

## ABSTRACT

---

### Which measure(s) of prosodic similarity as an evaluation of imitation?

Imitative proficiency across speakers is highly variable. Studies on imitation underline how difficult it is to find measurable cues to assess a successful imitation. In this exploratory study, *f0* contours collected from imitations tasks are tested in a double approach: objective measurements of prosodic similarity using two measures reported in the literature and perceptive evaluation by a panel of 15 naïve listeners. Our first results indicate a good correlation between the two approaches and they raise the question concerning the selection of the measurable factor assessing a successful imitation at a tonal level. Furthermore, these results underline an imitative proficiency's variability across speakers while opening perspectives in the domain of phonetic correction using the Verbo Tonal Method.

---

**MOTS-CLES** : parole, prosodie, imitation, mesures objectives, évaluations perceptives.

**KEYWORDS**: speech, prosody, imitation, objective measurements, perceptive evaluation.

---

## 1 Introduction

Les études sur l'imitation de la parole rapportent différents types de comportements imitatifs, comme la convergence (une adaptation mutuelle des interlocuteurs au fil de la conversation, Pardo,

2006), l'impersonnation (la tentative d'usurper la voix de l'autre, Révis, De Looze, & Giovanni, 2013), la simple imitation (Mixdorff, Cole, & Shattuck-Hufnagel, 2012) ou le *shadowing* (Dufour & Nguyen, 2013; Goldinger, 1998). Définir ces comportements en se basant sur des facteurs comme le contexte de production (Lewandowski, 2012) ou l'intention de l'imitateur (Donald, 1993) aboutit à les différencier. Pourtant, ils partagent une similitude définitoire majeure : pour être qualifiés de comportements imitatifs, la production du locuteur doit être perçue par un tiers comme similaire au modèle de l'imitation. Par conséquent, l'étude de l'imitation en parole vise à faire produire, puis à observer des changements comportementaux dans la manière de parler de locuteurs (naïfs ou experts), au niveau lexical ou phonétique. Pour ce qui est du niveau phonétique, deux questions ne cessent de représenter un défi méthodologique : quels sont les paramètres du signal sonore perçus puis imités en priorité par les locuteurs ; de quelle manière évaluer et comparer les paramètres choisis, entre leur modèle et leur(s) imitation(s) ? Choisir quel(s) paramètre(s) acoustique(s) mesurer et lier aux résultats d'évaluations perceptives de l'imitation demeure un choix crucial (Pardo, 2013).

Les imitateurs professionnels parviennent à ajuster globalement leur voix aux spécificités de leurs voix cibles, mais ils sont aussi capables d'imiter en reproduisant des variations instantanées de contours intonatifs ou de fréquence et durée des pauses (stratégies de synchronie) (Révis et al., 2013). En revanche, d'après ces derniers travaux, les locuteurs naïfs semblent limités à des stratégies d'ajustement global. Cette dernière remarque soulève une série de questions connexes, en lien avec les stratégies de synchronie : **(1)** jusqu'à quel point un locuteur naïf est-il capable de reproduire un patron prosodique perçu (variation instantanée) ; **(2)** est-il possible d'entraîner un locuteur à reproduire fidèlement des contours intonatifs ; **(3)** comment peut-on évaluer leur réussite –et leur échec– dans l'accomplissement de cette tâche

Les deux premières questions ont une pertinence certaine dans le domaine de l'enseignement de la prononciation à des locuteurs de langue seconde (L2), plus particulièrement en se plaçant dans le cadre théorique proposé par la Méthode Verbo-Tonale d'intégration phonétique (MVT). La MVT postule que les erreurs de prononciation en L2 seraient dues à un biais de perception de la L2. Afin de neutraliser les effets de ce biais, la MVT propose une rééducation perceptive, au moyen d'un ensemble de procédés correctifs où l'influence de la prosodie sur les segments phonétiques joue un rôle majeur. Un enseignant recourant à la MVT doit donc avoir une conscience et un contrôle prosodique efficace, notamment lorsqu'il doit produire des énoncés délexicalisés afin de faciliter la perception par l'apprenant des caractéristiques rythmiques et intonationnelles de la langue cible (Billières, Alazard, Astésano, & Nocaudie, 2013).

Intrinsèquement, une séquence de correction phonétique représente un cas typique d'interaction imitative. En effet, durant une interaction MVT, l'enseignant comme l'apprenant doivent imiter ou réitérer des sons de leur interlocuteur. L'apprenant doit répéter le modèle proposé par l'enseignant, ce qui peut conduire à questionner le lien entre perception et (re)production de la parole chez le sujet devenant bilingue. De son côté, l'enseignant doit produire systématiquement des patrons prosodiques phonologiquement cohérents, soulevant ainsi la question du contrôle de sa production, plus particulièrement au niveau mélodique.

Si les questions **(1)** et **(2)** peuvent être liées à la fois à l'enseignant et à l'apprenant, la présente étude se focalise sur la capacité de l'enseignant à reproduire systématiquement des patrons prosodiques. En effet, avant même de pouvoir tester la capacité de l'apprenant à (re)produire les paramètres phonétiques d'un énoncé, il faut s'assurer que l'enseignant même est capable d'imiter les éléments prosodiques saillants de la parole. Or, la pratique de la MVT implique la reproduction de paramètres

prosodiques de manière maîtrisée et la mise en valeur de certains événements pour faciliter à l'apprenant la perception des sons cibles. Ainsi, cette étude propose de tester en premier lieu la capacité des locuteurs de L1 à produire des imitations de contours prosodiques. Ce faisant, nous nous intéresserons plus particulièrement à notre question (3), à savoir : évaluer le degré de réussite d'une imitation. La méthode d'évaluation de la (dis)similarité prosodique pourrait conduire à la création d'un outil d'évaluation du contrôle prosodique de l'enseignant de L2, et ainsi être utile à leur formation dans le domaine de la correction phonétique.

La littérature sur l'imitation parolière en langue française est réduite, et peu d'études ont abordé plus spécifiquement la reproduction des indices prosodiques (voir cependant Michelas & Nguyen, 2011 à propos de l'accent initial). Cette communication est la poursuite d'une autre étude préliminaire qui décrivait la capacité de locuteurs à imiter les indices prosodiques de phrases contrôlées, au cours de trois tâches d'imitation, d'une simple répétition à une exagération (Nocaudie & Astésano, 2012).

## 2 Matériel linguistique : un corpus d'imitation(s)

Notre corpus d'imitation est issu d'un corpus de phrases (*Corpus d'Edimbourg : CE*) présentant une ambiguïté syntaxique qui peut être résolue par la production des indices prosodiques pertinents. L'ambiguïté syntaxique dérive de la manipulation de la portée de l'adjectif, comme dans « les gants et les bas lisses », où l'adjectif (A) « lisses » qualifie alternativement le second nom « bas » uniquement ([les gants][et les bas lisses]) (*Condition 1 : Cond-1*) ou bien les deux noms ([les gants et les bas][lisses]) (*Condition 2 : Cond-2*). La longueur des noms et des adjectifs varie de une à quatre syllabes. Le contrôle de l'ambiguïté syntaxique et de la longueur des constituants nous permet d'observer les indices prosodiques (proéminences, frontières, tons, pauses...) utilisés dans la linéarisation syntaxique des énoncés oraux (voir Astésano, Bard, & Turk, (2007) pour le détail de la constitution du *CE*).

Du *CE*, 16 énoncés prononcés par une locutrice ont été sélectionnés comme stimuli auditifs pour la constitution du corpus *Imitation (CI)*. Ces phrases comportaient uniquement deux longueurs de noms (noms tri- et quadrisyllabiques) combinées à des adjectifs de une à quatre syllabes, prononcées dans les deux conditions syntaxiques (*Cond-1 & Cond-2*). 8 locuteurs naïfs ont dit ces phrases au fil de trois tâches différentes : a) une simple répétition (*REP*), b) une imitation (*IMI*), c) une exagération des phrases de la locutrice (*EXA*). Les données de 2 locuteurs ont été exclues du *CI* en raisons de facteurs physiologiques ayant eu une incidence sur leur voix (stress induit par la situation expérimentale, timbre éraillé). Durant chaque tâche, les imitateurs répétaient chaque phrase 3 fois, dans un ordre aléatoire, nous permettant d'obtenir un total de 864 phrases (16 énoncés \* 2 conditions syntaxiques \* 3 occurrences \* 3 tâches \* 6 imitateurs). Afin d'évaluer la capacité implicite des locuteurs à imiter la parole, l'attention des locuteurs n'était pas focalisée sur le fait d'imiter durant la tâche a). Il leur était simplement demandé de « dire la phrase entendue en préservant sa structure ». Durant les tâches b) & c), il leur était demandé explicitement de produire une imitation ou une exagération des énoncés entendus. Cependant, l'expérimentateur n'a pas dirigé l'attention des imitateurs sur les indices prosodiques.

La présente étude vise à comparer des données perceptives (Test AX) et objectives (issues d'un algorithme) obtenues à partir d'un extrait du *CI*. Nous avons sélectionné un sous-corpus de 4 énoncés de 2 longueurs différentes pour tester la robustesse de notre algorithme de mesure de la similarité prosodique. En effet, toutes les phrases sont prises dans la condition *Cond-1* car la désambiguïtation syntaxique est normalement marquée par une pause silencieuse entre le premier et

le second nom. Comme l'algorithme a une nette tendance à aligner les silences lorsqu'il aligne les contours, l'absence de pause dans l'imitation devrait gêner l'alignement avec la référence et se traduire théoriquement par un score de similarité moins élevé.

Les tests ont été menés sur la production de 4 imitateurs (Sp1, Sp3, Sp5 & Sp7), qui étaient appariés pour prononcer certains énoncés. Sp1 (femme) & Sp5 (femme) ont dit « *Les bagatelles et les balivernes sottes* » et « *Les bonimenteurs et les baratineurs fades* » ; Sp3 (homme) & Sp7 (femme) ont prononcé « *Les bagatelles et les balivernes saugrenues* » et « *Les bonimenteurs et les baratineurs fabuleux* ». Nos résultats ont été calculés sur 18 énoncés par locuteur, produisant alors un nombre total de 72 couples de phrases (modèle vs. imitation) : [(2 énoncés \* 3 répétitions \* 3 tâches) \* 4 locuteurs].

### 3 Méthode : mesures objectives & évaluations perceptives d'imitations prosodiques

Un problème constant dans l'évaluation de l'imitation en parole réside dans la relative absence de congruence entre la multitude de paramètres acoustiques pouvant être mesurés, alors que ces derniers divergent ou convergent avec le modèle. La mélodie, et son corrélat physique, la  $f_0$  ont été décrits comme une cible de choix pour l'imitateur (Révis et al., 2013). Par ailleurs, les procédés correctifs de la MVT reposent largement sur la manipulation des contours mélodiques. Ainsi, notre méthode se concentre principalement sur la mesure de la distance physique de contours de  $f_0$  appariés (modèle vs. imitation) d'une part et sur l'évaluation perceptive de leur ressemblance d'autre part.

#### 3.1 Dynamic Time Warping (DTW) & (dis)similarité prosodique

Mesurer objectivement l'imitation d'un contour mélodique équivaut à trouver s'il y a une distance physique entre le contour modèle et sa reproduction, *i.e.* connaître le degré de correspondance entre les deux formes de contours intonatifs.

Ceci étant dit, comparer des formes impose une normalisation tonale et un alignement temporel des pics et des creux décrits par les contours de  $f_0$  (DTW). Cette méthode d'interpolation non linéaire, qui force l'alignement entre le modèle et sa reproduction, améliorerait les scores de corrélation entre les formes, notamment si les contours intonatifs sont fonctionnellement similaires, *i.e.* s'ils partagent le même patron accentuel (Rilliard *et al.*, 2011). La distance entre les paires de contours intonatifs a été calculée au moyen de deux mesures similaires à celles proposées par Hermes (1998) où :  $w(t)$  correspond au décours temporel du facteur de poids (la somme des *subharmonic sumspectrum* du contour imité),  $W$  est l'intégrale du contour de 0 à  $T$  (avec  $T$ , la durée totale du contour) et  $f_1$  &  $f_2$ , la paire de contours testée par l'algorithme.

Pour ces travaux, la procédure de normalisation de la  $f_0$  choisie diffère de celle d'Hermes. En effet, nous avons divisé chaque valeur de  $f_0$  par la valeur maximale de  $f_0$  de l'énoncé ( $f_1 = 1/p_{1max}$ ). Cette procédure de normalisation permet de comparer plus facilement les imitateurs masculins et féminins en ramenant la variation des courbes de  $f_0$  dans des valeurs comprises entre 0 et 1. Par ailleurs, cette procédure de normalisation facilite le travail de la partie DTW de l'algorithme. Les valeurs de  $f_0$  ont été extraites à l'aide du logiciel PRAAT (Boersma, 2001) et le taux d'échantillonnage était d'une valeur par milliseconde (Rilliard, Allauzen, & de Mareüil, 2011).

Après normalisation et DTW, nous avons calculé la différence de la moyenne des moindres carrés (ci-après,  $L_2$ ) de chaque paire de contours comme suit :

$$L_2 = \left\{ \frac{1}{W} \int_0^T w(t) |f_1(t) - f_2(t)|^2 dt \right\} \quad (1)$$

Le coefficient de corrélation  $r$  entre deux contours  $f_1$  et  $f_2$  ont ensuite été calculés de la manière suivante :

$$r = \frac{\frac{1}{W} \int_0^T w(t) f_1(t) f_2(t) dt}{\sqrt{\left\{ \frac{1}{W} \int_0^T w(t) |f_1(t)|^2 dt \frac{1}{W} \int_0^T w(t) |f_2(t)|^2 dt \right\}}} \quad (2)$$

Afin de pouvoir comparer les coefficients de corrélation, Hermes propose de transformer  $r$  en  $Z$  de Fisher (ci-après,  $Zr$ ), que nous calculons ainsi :

$$Z_{f_1 f_2} = \frac{1}{2} \ln \frac{1+r_{f_1 f_2}}{1-r_{f_1 f_2}} \quad (3)$$

$L_2$  mesure les changements rapides de  $f_0$ , tandis que  $Zr$  est une mesure holistique de la proximité de la forme de deux contours. Il est intéressant de tester leur complémentarité vis-à-vis des jugements perceptifs que nous recueillons par ailleurs.

Finalement, chaque phrase reproduite a reçu un rang en fonction de ses scores  $L_2$  et  $Zr$ .  $L_2$  est une mesure de dissimilarité (soit, plus  $L_2$  a une valeur élevée, plus grande est la dissimilarité entre le modèle et sa reproduction). La phrase avec le  $L_2$  le plus bas a donc été classée comme rang 1, celle avec le  $L_2$  le plus haut, rang 72.  $Zr$  est une mesure de similarité (plus  $Zr$  est haut, plus la similarité entre les deux contours est grande). La phrase avec le plus haut  $Zr$  a donc reçu le rang 1, celle avec le plus bas, le rang 72. Ainsi, nous avons obtenu 2 classements différents (en fonction de  $L_2$  ou de  $Zr$ ), qui seront comparés au classement dérivé des résultats des évaluations perceptives.

### 3.2 Test AX de jugement de la similarité

Comme le note Pardo (2013), l'imitation en parole devrait être évaluée objectivement et subjectivement, *i.e.* physiquement et perceptivement. A cette fin, nous avons complété les mesures objectives décrites ci-dessus avec des données issues d'une tâche AX de jugement de la similarité. Cette tâche permet d'obtenir une évaluation holistique de chaque phrase imitée (X) comparée à son modèle (A). 15 auditeurs naïfs ont pris part à la tâche. Tous étaient de L1 française (âge : 25-32) et ne présentaient pas de trouble de la parole ou de l'audition.

Il était demandé aux locuteurs de noter la ressemblance de X avec A en termes de musicalité de la parole (rythme, variation tonale). Le test AX a été passé au moyen du logiciel Lancelot (environnement HTML de PERCEVAL (André, Ghio, Cavé, & Teston, 2003)). Les paires de phrases étaient randomisées par le logiciel, et présentées en modalité auditive dans des écouteurs de qualité professionnelle. Les auditeurs pouvaient écouter chaque couple de phrase jusqu'à cinq fois avant de leur attribuer une note sur une échelle Likert allant de 1 (moins similaire) à 5 (très similaire) en touchant l'écran tactile de l'ordinateur de passation.

Pour chaque phrase X, la moyenne des 15 scores obtenus a été calculée. La phrase obtenant la moyenne la plus élevée a obtenu le rang 1 du classement AX, etc. En cas d'égalité de score entre deux phrases X ou plus, un rang égal à la moyenne des rangs qu'elles devraient occuper dans le classement a été attribué à chacune d'entre elles (par exemple : 7 ; 8 ; 9  $\rightarrow$  8 ; 8 ; 8).

## 4 Résultats

Nous décrivons tout d'abord les résultats comparant les classements obtenus à partir des scores objectifs ( $L_2$  &  $Zr$ ) et du score perceptif (AX) pour tous les locuteurs. La Figure 1 montre la distribution des rangs pour les trois types de scores, qui donneront par la suite lieu à un calcul de corrélation. Sur ce diagramme, le point indique le score moyen obtenu par le locuteur et les boîtes montrent la distribution interquartile des rangs de chacune des 18 phrases prononcées par locuteur, parmi l'ensemble de 72 phrases évaluées. Les barres de confiance montrent les valeurs minimales et maximales de rangs obtenus par chaque locuteur, en fonction du classement. Nous rappelons que le rang 1 montre la production qui a obtenu le meilleur jugement de similarité.

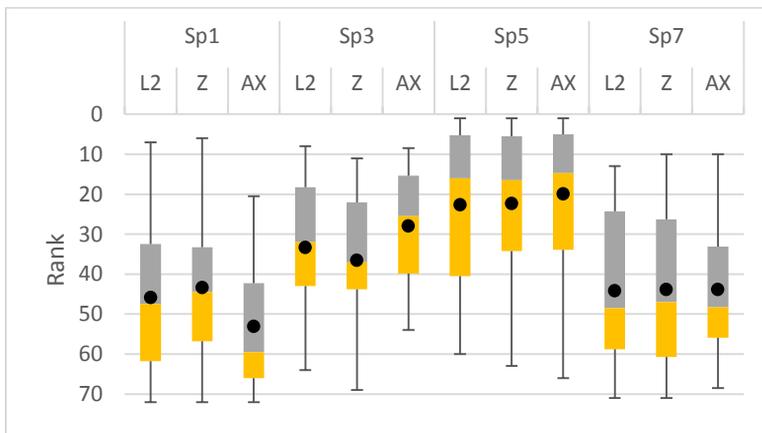


Figure 1 : Distribution des rangs (de 1 à 72, en ordonnées) obtenus par chaque locuteur (Sp1, Sp3 ; Sp5 & Sp7) en fonction des types de scores ( $L_2$ ,  $Zr$  & AX). Les points représentent le rang moyen de chaque locuteur pour les 18 énoncés imités.

À la vue de cette distribution des rangs obtenus par chaque locuteur, il semble raisonnable de classer Sp5 comme le locuteur le plus performant dans les tâches d'imitation (Rang moyens :  $L_2 = 22,76$  ;  $Zr = 22,28$  ;  $AX = 19,88$ ), suivi clairement par Sp3 ( $L_2 = 33,33$  ;  $Zr = 36,56$  ;  $AX = 27,92$ ). Les rangs des classements objectifs de Sp7 ( $L_2 = 44,17$  ;  $Zr = 43,83$ ) et de Sp1 ( $L_2 = 45,83$  ;  $Zr = 43,33$ ) sont très proches, mais leur rangs du classement AX (Respectivement :  $AX(Sp7) = 43,82$  ;  $AX(Sp1) = 53,06$ ) peut refléter la plus grande dispersion de leurs rangs dans le quartile inférieur : le meilleur rang obtenu par Sp1 est meilleur que celui de Sp7, mais cette valeur isolée peut fausser le calcul de la moyenne de leurs rangs. De fait, Sp7 a obtenu un plus grand nombre de bons rangs que Sp1 au fil des procédures d'évaluation, comme le résume le diagramme.

Le calcul de la corrélation a été fait au moyen de Real Statistics pour Excel (Zaiontz, 2015). Afin de comparer les données objectives et subjectives, nous proposons d'utiliser le  $r_s$  de Spearman, qui est un indice de corrélation entre des données *ordonnées* (rangs des classements). Les tests bilatéraux de Spearman menés sur nos données montrent des corrélations positives :

- $Zr$  vs. AX  $\rightarrow r_s = .554, p < .0001, t(71) = 5.562$
- $L_2$  vs. AX  $\rightarrow r_s = .589, p < .0001, t(71) = 6,092$

Les deux indices de corrélation dépassent les valeurs critiques de  $r_s$  admises pour  $N = 72$  ( $r_{s-crit} = .382$ ,  $t_{-crit} = 3.43$ ). Ainsi, la relation linéaire entre les classements objectifs et perceptifs semble robuste.

La Figure 2A montre la relation entre  $L_2$  et  $Z_r$  pour les 72 phrases testées. Les points dans le coin en bas à droite sont les phrases qui ont été jugées très similaires à leur modèle. Les résultats donnés par l'algorithme soulignent la différence de performance en imitation entre les différents locuteurs. La Figure 2B illustre la différence de performance entre les deux locuteurs qui ont été respectivement classés le moins (Sp1) et le plus (Sp5) performants au fil des tâches, d'après les résultats de l'algorithme et du panel d'auditeurs.

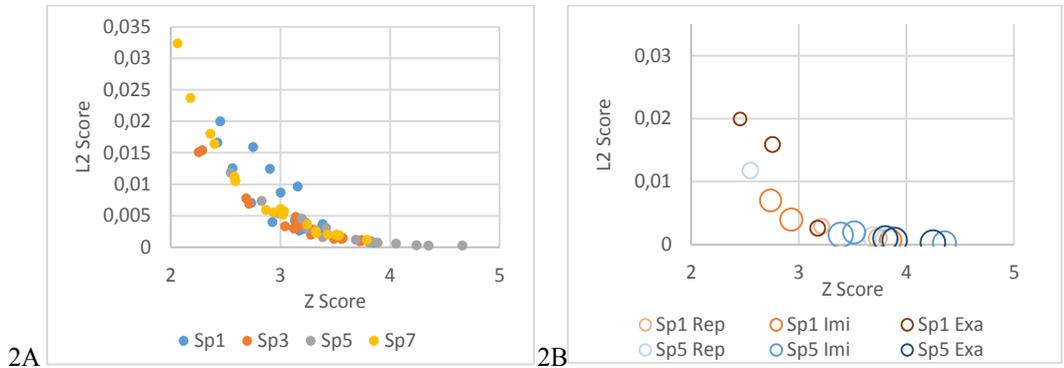


Figure 2 (A & B) : La Figure 2A montre les scores  $L_2$  (axe  $y$ ) &  $Z_r$  (axe  $x$ ) de 71 phrases. La Figure 2B illustre en trois dimensions l'évaluation de la performance de 2 locuteurs imitant la phrase « Les bagatelles et les balivernes sottes » au cours de REP, IMI & EXA. Les scores  $Z_r$  et  $L_2$  sont respectivement sur les axes des  $x$  et des  $y$  (plus  $L_2$  est haut, moins la paire est similaire ; plus  $Z_r$  est haut, plus la paire est similaire), la taille des points représente le score moyen au test AX. Un point d'exception a été retiré de ces représentations (Sp1 Rep  $L_2=0.0615$ ,  $Z_r=1.6903$ , AX =1.4) afin d'améliorer la lisibilité de ces représentations.

Nous prévoyions que les imitations les plus conscientes (réalisées durant les tâches *IMI* & *EXA*) présenteraient une reproduction plus précise des indices prosodiques. Si cette prédiction se retrouve dans les résultats de Sp5 dont la performance, de mieux en mieux notée au fil des tâches, semble montrer un bon contrôle prosodique ; notre prévision se trouve mise à mal par la performance de Sp1 dont certaines phrases produites en *REP* obtiennent de meilleures notes qu'en *IMI* ou en *EXA*. Ainsi, nos résultats mettent en relief une grande variabilité interlocuteur en termes de contrôle prosodique au cours des tâches.

## 5 Discussion & Conclusion

Cette étude visait à tester méthodologiquement une approche double d'évaluation/mesure de l'imitation des formes prosodiques. A terme, nous souhaitons développer un algorithme suffisamment robuste pour être implémenté dans un outil d'entraînement des enseignants qui leur permettrait d'évaluer la précision de leur performance prosodique.

Précédemment, le DTW, dont le terme  $d$  indique le coût d'alignement entre deux contours, a été utilisé par Kim (2012) en tant que mesure de la convergence phonétique. Dans notre cadre, le DTW

est principalement utilisé comme méthode d'interpolation (Rilliard et al., 2011) préliminaire au calcul de deux mesures de la similarité prosodique (initialement rapportées par Hermes, 1998). D'après ce dernier,  $L_2$  mesure la distance perceptive entre deux contours, où une distance élevée a un facteur de poids quadratiquement plus élevé.  $Zr$  exprime une distance globale entre deux formes de contours, *i.e.* sa valeur indique le coût de la transformation d'un contour en un autre. Etant données leurs natures différentes, il était intéressant de les corrélérer tous deux aux résultats des évaluations perceptives.

Ces deux méthodes de mesure de la similarité prosodique ont obtenu une bonne corrélation avec les résultats du test AX conduits auprès de 15 auditeurs naïfs : les bonnes et les mauvaises imitations ont été repérées par l'algorithme comme par les auditeurs. La différence de corrélation entre AX et  $L_2/Zr$  pourrait refléter la nature différente des mesures objectives. Cela étant dit, ces résultats ouvrent des perspectives intéressantes dans l'évaluation automatique de l'imitation au niveau prosodique. La question des mesures automatiques reste cependant ouverte : une étude sur plus de sujets et plus de stimuli pourrait nous renseigner sur la nécessité de conserver  $L_2$  et/ou  $Zr$ . De même, de plus amples investigations seront nécessaires à la détermination d'une valeur seuil pour  $L_2/Zr$ , au-delà de laquelle le résultat d'un test perceptif de jugement de la similarité prosodique pourrait être estimé avec une précision suffisante.

Ces premiers résultats, encourageants, nous poussent à étendre cette approche dans de multiples directions : (1) plus d'énoncés du *CI* seront notés, objectivement et subjectivement ; (2) un corpus de phrases et de leur reproduction délexicalisée par des locuteurs sera constitué afin de tester la robustesse des mesures.

Par ailleurs, il pourrait être pertinent d'envisager l'utilisation d'autres méthodes de transformation que le DTW. Ce dernier requiert un grand nombre de valeurs de  $f\theta$ , conduisant à un temps de calcul relativement long. Parmi les méthodes de comparaison de formes passées en revue par (Veltkamp, 2001), la fonction de cumulation des angles semble pouvoir être appliquée à des contours de  $f\theta$  stylisés à partir de quelques points remarquables. Cette approche permettrait de prendre en compte explicitement certains détails des patrons prosodiques, comme les durées et les pentes et ainsi améliorer la description de ce qu'est une imitation réussie. Enfin, le raffinement de ces mesures, devrait à terme nous conduire à limiter drastiquement le recours aux tests perceptifs, en sélectionnant les mesures objectives corrélant au mieux avec des résultats perceptifs extensifs.

Les tâches accomplies pour recueillir le *CI* visaient à souligner la capacité d'imitation de locuteurs naïfs, soit, un comportement approchant celui d'un enseignant de langue étrangère sans contrôle prosodique particulier, lors de tentatives de correction phonétique. Pour certains locuteurs (par exemple, Sp1), les scores objectifs fournissent une aide au diagnostic du niveau de contrôle et de conscience prosodique, et dans une certaine mesure, du talent phonétique. Lewandowski (2012) rapporte en effet la complexité à déterminer le talent phonétique d'une personne. Ce type de mesures automatique pourrait donc être utilisé comme un indice pour détecter une composante du talent phonétique, plus précisément au niveau prosodique.

Finalement, nos perspectives de recherches se concentreront sur certains aspects spécifiques de l'entraînement à la MVT, plus particulièrement, sur la correction des indices prosodiques reproduits. A court terme, nous espérons construire une interface permettant à l'enseignant d'améliorer ses compétences d'utilisation des procédés de correction de la MVT, dans ce cas, la production de phrases délexicalisées servant à porter l'attention de l'apprenant sur la syllabification et le rythme de la langue en cours d'apprentissage.

## Références

- ANDRE, C., GHIO, A., CAVE, C., & TESTON, B. (2003). PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. In *Proceedings of XVth ICPHS* (p. 1421-1424). Barcelone, Espagne.
- ASTÉSANO, C., BARD, E. G., & TURK, A. (2007). Structural influences on Initial Accent placement in French. *Language and speech*, 50(3), 423-446.
- BILLIERES, M., ALAZARD, C., ASTESANO, C., & NOCAUDIE, O. (2013). Phonétique corrective en FLE : Méthode Verbo-Tonale. <http://w3.uohprod.univ-tlse2.fr/UOH-PHONETIQUE-FLE/>
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5:9/10, 341-345.
- DONALD, M. (1993). *Origins of the Modern Mind - Three Stages in the Evolution of Culture & Cognition* (Reprint). Cambridge, Mass.: Harvard University Press.
- DUFOUR, S., & NGUYEN, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, 4.
- GOLDINGER, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-279.
- HERMES, D. J. (1998). Measuring the Perceptual Similarity of Pitch Contours. *Journal of Speech, Language and Hearing Research*, 41, 73-82.
- KIM, M. (2012). *Phonetic accommodation after auditory exposure to native and nonnative speech*. NORTHWESTERN UNIVERSITY.
- LEWANDOWSKI, N. (2012). *Talent in non-native phonetic convergence*. Universität Stuttgart, Stuttgart.
- MICHELAS, A., & NGUYEN, N. (2011). Uncovering the Effect of Imitation on Tonal Patterns of French Accentual Phrases. In *INTERSPEECH* (p. 973-976).
- MIXDORFF, H., COLE, J., & SHATTUCK-HUFNAGEL, S. (2012). Prosodic Similarity—Evidence from an Imitation Study. In *Speech Prosody 2012*.
- NOCAUDIE, O., & ASTÉSANO, C. (2012). Prosodic structuring imitation in French L1 context—A first step towards correcting phonetic-prosodic features in L2 French. In *Proceedings of ISICS*. Aix-en-Provence.
- PARDO, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382-93.
- PARDO, J. S. (2013). Reconciling diverse findings in studies of phonetic convergence. In *Proceedings of Meetings on Acoustics* (Vol. 19, p. 060140). Acoustical Society of America.
- REVIS, J., DE LOOZE, C., & GIOVANNI, A. (2013). Vocal Flexibility and Prosodic Strategies in a Professional Impersonator. *Journal of Voice*, 27(4), 524.e23-524.e31.
- RILLIARD, A., ALLAUZEN, A., & DE MAREÛIL, P. B. (2011). Using Dynamic Time Warping to Compute Prosodic Similarity Measures. In *INTERSPEECH* (p. 2021-2024).
- VELTKAMP, R. C. (2001). Shape matching: similarity measures and algorithms (p. 188-197). IEEE Computer Society.
- ZAIONTZ, C. (2015). *Real Statistics Using Excel*. Consulté à l'adresse [www.real-statistics.com](http://www.real-statistics.com)