People with Parkinson's disease exhibit phonetic flexibility

Véronique Delvaux[1,2] *, Bernard Harmegnies*[1]*, Kathy Huet*[1]*, Myriam Piccaluga*[1] *Virginie Roland*[1]*, Clémence Verhaegen*[1]

[1]Institut de Recherche en Sciences et Technologies du Langage, UMONS, Belgique
[2]Fonds de la Recherche Scientifique, FNRS, Belgique
veronique.delvaux@umons.ac.be

Abstract

Phonetic flexibility is defined as the ability of the speaker-listener to adapt his speech behavior to the internal and external constraints weighing on the communication situation. This exploratory study deals with phonetic flexibility in elderly people with Parkinson's disease (PD). Twelve participants aged between 57 and 79, half of whom had PD, produced the 25 sentences of the corpus under 6 successive experimental conditions: (i) reading; (ii) repetition of an oral model (= target voice); (iii) interactive game; (iv) repetition of an oral pattern; (v) repetition with explicit instructions to imitate the oral model; (vi) repetition with explicit instructions to inhibit imitation of the oral model. The results show that the participants, regardless of whether or not they belonged to the PD group, massively showed phonetic flexibility. Indeed, the measured pitches and durations systematically moved away from those measured in reading to approach those of the target, depending on the experimental condition, namely: moderately in repetition (phonetic convergence), largely in imitation, almost not in inhibition.

Index Terms: Phonetic flexibility, phonetic convergence, repetition, imitation, Parkinson's disease, dysarthria

1. Introduction

Phonetic flexibility is defined as the ability of the speaker-listener to adapt his/her speech behavior to the internal and external constraints of the communication situation. Speech is thus considered as a complex behavior supported by multiple, complex and interactive cognitive processes, under the control of the executive functions.

In the linguistic tradition, it is generally considered that the phonic form of messages is determined by linguistic constraints and rules, but also depends on a large number of so-called paralinguistic factors (attitudes, emotions, dialectal or sociocultural origin of the speaker, etc.), or even extra-linguistic factors (sex, age, physical attributes of the speaker). The second half of the 20th century has documented many cases of phonetic variation related to the dialectal or sociolinguistic origin of the speakers. The last twenty years have seen many advances in our understanding of the effect on speech and voice of many other, previously little explored, factors, e.g., acoustic correlates of body size and mass (Pisanski et al., 2014), vocal expression of emotion (Grandjean and Baenziger, 2014), prosodic characterization of attitudinal factors such as irony (Rouas et al., 2019), effects of stress and cognitive load (Huttunen et al., 2011), etc.

In traditional linguistics, since the focus is on the form of the message, it is considered to be the resultant - unique? - of the combined action of the linguistic, paralinguistic and extralinguistic factors mentioned above. We propose to reverse the perspective and to consider speech from the point of view of cognitive sciences centered on the human subject, the speaker-listener. Each speech production is then the unique and irreducible product of an interaction between human subjects in a specific communication situation. The different factors determine a framework of constraints - the field of possibilities - but do not allow the prediction of a unique phonic form. The individual shows flexibility, adaptability and phonetic control. According to his needs, expectations, preferred strategies, which can themselves evolve in relation to the situation, the speaker constantly adapts, in real time, the phonetic form of his speech productions. We therefore assume that part of the variability in speech production, both inter-individual and intra-individual, is in fact the product of the control exercised by each speaker on his phonetic behavior in order to respond to the specific internal and external constraints imposed by the language situation and all its components (transactional model of the environment/behavior relationship: Sameroff, 2009).

Recent work has demonstrated the existence of important resources in speakers in terms of what we call "phonetic flexibility". This is evidenced by the numerous studies on imitation and vocal disguise abilities (Revis et al., 2013). Moreover, it is now accepted that this flexibility potential is used daily by speakers during language interactions (cf. the phonetic convergence mechanism typically occurring between two interlocutors: Pardo et al., 2018), as well as all the phenomena of adaptation to the interlocutor: 'child-directed speech', 'foreigner-directed' speech, speech addressed to the hearing impaired, etc.; for a review, see Cooke et al., 2014). More generally, a large literature concerns adjustments to speech productions in response to constraints on the communication situation, both 'external' constraints (e.g., speech in noise, Garnier and Heinrich, 2014), and constraints internalized by the speaker (the traditional hypo- vs. hyper-articulation continuum: Moon and Lindblom, 1994; Smiljanic and Bradlow, 2008). Finally, recent experiments using various speech perturbation paradigms have shown the extent to which speakers are able to adjust their motor programs in real time in order to compensate, totally or partially, for the effects of an articulatory (lip tube, bite block) or auditory (altered feedback) perturbation, and thus to make the best use of speech motor control mechanisms in order to approach their target in sensory-motor and/or acoustic-auditory domains in spite of everything (in adults as well as in children: Menard et al., 2016).

These adjustment/adaptation mechanisms of the "all-round" speaker constitute in fact as many resources accessible to the individual confronted with a set of disturbances due to the appearance of a pathology affecting speech. Beyond the effects of the pathology itself, of the drug treatment, as well as their interactions, it is commonly admitted that patients set up specific compensatory strategies in order to maintain an adapted production of their speech sounds. These compensatory strategies, to be effective, must be based on a solid foundation of phonetic flexibility, as well as on other elements such as preserved access to cognitive resources (elderly and/or neurodegenerative patients) and motivation (personal, professional) to maintain rich and nuanced spoken communication with the environment.

Previous work in our laboratory has investigated phonetic flexibility abilities via a phonetic learning paradigm of aspirated plosives in French-speaking speakers over 65 years of age (Delvaux et al., 2015), as well as in individuals with speech aphasia (Verhagen et al., 2020). In the study reported here, we explore phonetic flexibility ability in elderly individuals with and without PD via a repetition/imitation/inhibition paradigm (detailed below). Parkinson's disease (PD) is a neurodegenerative disease responsible for progressive disorders of the motor sphere. Among the associated disorders is dysarhtria, a generalized disorder of motor execution of speech, involving difficulty in initiation, lack of coordination or slowness in the execution of articulatory gestures, which typically results in dysphonia, articulatory imprecision and prosodic insufficiency (monotonous, monotonous voice, timing abnormalities), associated or not with respiratory problems (Rusz et al., 2011). At an early stage of the disease, the atypicalities observed in the speech may be consecutive to the disorder itself (or even to the drug treatment), but also to compensatory strategies put in place by the patient to cope with it (e.g. maintaining speech rate at the expense of articulatory precision).

The principle of the experimental paradigm used here is to induce in the participants controlled variations in speech rate and pitch register, linked to instructions associated with a succession of experimental tasks or situations involving spoken communication. In addition to the clinical perspectives, this research aims at better understanding the cognitive and linguistic mechanisms that support the production of speech, conceived as a strategic behavior.

2. Material and methods

2.1. Participants

Twelve participants, six men and six women between the ages of 59 and 79, took part in the study. Half of them had Parkinson's disease. Relevant biographical characteristics are collected in Table 1.

Table 1: *This is an example of a table.*

Sujet	Groupe
S01	Ctrl
S02	Ctrl
S03	Ctrl
S04	Ctrl
S05	Ctrl
S11	Ctrl
S06	MP

2.2. Corpus

The corpus consists of 25 sentences, 5 initial sentences (produced by a 22-year-old native French speaker) each having undergone 5 types of modifications. These modifications were performed via the "Manipulate" procedure in Praat and concern either the total duration (75%, 100%, 125% of the initial duration) of the sentence, or its average pitch (75%, 100% or 125% of the initial average pitch). The initial phrase is thus coded D100P100 and the four other versions : D75P100, D125P100, D100P75, D100P125. The sentences are: "Lalie sells lilacs", "The wolf lives in the woods", "Is he a daddy or a grandpa?", "Does he have a beard or a moustache?", "Does he have a big mouth or a small mouth?" The sentences in the corpus vary in length (6 to 11 syllables), syllabic complexity or phonological content (e.g., the first two sentences are totally voiced), prosodic pattern, etc.

2.3. Experimental paradigm

The experimental paradigm consists of six successive steps, completed in a single experimental session. At each stage of the paradigm (except (iii)), the participant produces the 25 sentences of the corpus according to a specific task: (i) reading; (ii) repetition of an oral model (the "target" voice); (iii) interactive play; (iv) repetition of an oral model; (v) repetition with explicit instruction to imitate the oral model; (vi) repetition with explicit instruction to inhibit as much as possible the imitation of the oral model. The 25 sentences are presented in the same pseudorandom order at each stage. The interactive game is a slightly adapted version of the board game "Who is this?", so only certain sentences are likely to be produced, without guarantee. The games (at least four games per participant) are all played between the participant and one and the same interlocutor, the one to whom the target voice belongs. The speech productions during the game will not be analyzed here.

2.4. Recordings and measurements

Acoustic data were recorded using a cell phone (iPhone 8) placed 30 cm from the participant's mouth on a separate stand. The recordings were converted to .wav format (44100Hz, mono) and processed in Praat. A dedicated script allowed: the alignment between the produced sentences (transcriptions) and the speech signal as well as the taking of measurements under the supervision of the user. The measures are: the total duration of the sentence, then converted into speech rate (number of syllables per second), as well as the fundamental frequency (Hz) calculated every 5 ms, from which we extracted the median value, expressed in z-score. For each utterance of each participant, we thus have two indicators, which will form the main dependent variables of the statistical analysis: the median pitch (z-score) and the rate (syll/s). Moreover, for tasks (ii), (iv), (v), (vi), we transformed these two indicators into a "convergence index". The principle is to express (in %) the proportion of the "path traveled" to the "distance to go" to the target. The absolute value (e.g., the average measured pitch for the repetition of the phrase "Lalie sells lilacs" D100P125) is expressed relatively, to the target:

((MeasuredPitch-PitchBaseline)/(TargetPitch PitchBaseline))*100

Thus, for a repetition of D100P125: (130Hz-110Hz) / [(125%*172Hz)-110Hz] = 19% convergence. The baseline retained is the average of the values measured in this participant during the reading task. This index can indicate a partial convergence (e.g. 25%) or almost total convergence (90-100%), or even an overshoot of the target (150%) or a divergence (negative value, e.g. -40%).

3. Results

Two repeated measures analyses of variance (ANOVA) were first performed, one with speech rate (syll/s) as the dependent variable, the other with pitch (z-score). The within-subjects factor is the task (repetition before the game; repetition after the game; imitation; inhibition), the between-subjects factor concerns the characteristics of the stimuli (respectively, for the rate: D75, D100, D125; for the pitch: P75, P100 and P125). In these analyses, the participant is considered as a random factor and the group (control vs. PM) is not taken into account. This is therefore an initial exploration of the data, all subjects combined.

The analyses reveal a significant effect of task and stimulus characteristics on flow and pitch, as well as a significant interaction of these two factors (Flow: $F(4,40.003)=30.255$, $p<.001$; Pitch: $F(4,40.081)=10.272$, $p<.001$), illustrated in Fig.1. In the repetition tasks, speakers tend to "follow" the model, producing sentences with a higher rate/slow in response to a higher rate/slow (similarly for pitch). In the imitation task, the differences observed in production as a function of stimulus characteristics are exacerbated (the slopes are steeper in Fig.1). In the inhibition task, the line is almost horizontal, which means that speakers globally manage to maintain a given level of rate or pitch, independently of those characterizing the voice that pronounces the sentences to be repeated.

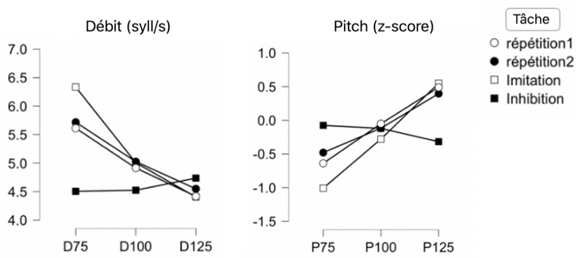

Figure 1: *Left: speech rate (syll/s) as a function of task and stimulus rate (D75: reduced duration = accelerated rate, D100: median duration/rate, D125: extended duration = slowed rate). Right: pitch (z-score) as a function of task and stimulus pitch (P75: lowered pitch, P100: median pitch, P125: raised pitch). All participants combined.*

In a second step, the effect of group (participants with or without Parkinson's disease) was studied. A multivariate analysis of variance was performed with the convergence rates calculated for flow and pitch as dependent variables and the task (repetition: includes pre- and post-game repetition tasks; imitation; inhibition) and group (control vs. PD) as independent variables.

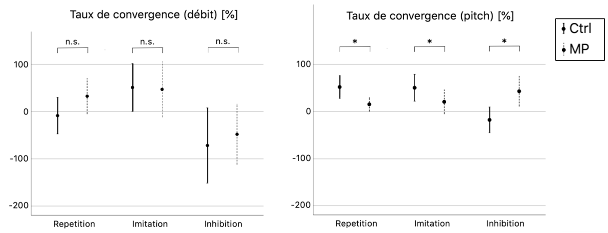

FIGURE 2. Convergence rate (%) as a function of task (repetition, imitation, inhibition) and group (Ctrl: controls; MP: Parkinson's disease). Left: Speech rate, right: Pitch.

Regarding the rate of convergence on speech rate, the analysis reveals only a significant effect of task ($F(2,1045)=6.713$, $p=.001$). There was no group effect, i.e. subjects with Parkinson's disease behaved like control subjects for speech rate (Fig.2, left): there was low convergence in repetition, high convergence (around 50% on average) in the imitation task and high divergence rates in inhibition. Regarding the convergence rate on the pitch, only the Group*Task interaction is significant ($F(2,1045)=7.939$, $p<.001$). Fig.2 (right) illustrates this interaction: speakers in the control group converge on the pitch of the target voice, both in repetition and in imitation, and manage to counteract this tendency in inhibition. On the other hand, speakers with Parkinson's disease are unable to follow the given instructions, especially in the inhibition task: they come (slightly) closer to the pitch of the target voice in all tasks.

Note that when we add the Sex factor to the independent variables in the statistical analysis, we observe a significant triple interaction Sex*Group*Task for the rate of convergence on the pitch ($F(2,1039)=7.915$, $p<.001$). This interaction reflects the fact that the trend shown in Fig.2 (right) concerns exclusively female subjects. Men, with or without PD, do not converge on pitch, regardless of the experimental task.

4. Discussion

The aim of this study was to explore phonetic flexibility in French-speaking participants over 60 years of age, half of whom have Parkinson's disease. The paradigm used allows to study the fluctuations of phonetic parameters of speech productions in relation to the production conditions. On the one hand, the instructions vary according to the task: simple repetition, explicit imitation, repetition without imitation, of the model voice. On the other hand, the sentences to be produced were manipulated in such a way as to obtain 5 versions of the same sentence, differing only in the rate of speech or the register of f0.

First of all, the results show that the subjects of the study are globally capable of phonetic flexibility. Their productions vary in the expected direction. When asked to simply repeat the sentences produced by a target voice, participants tend to imitate the speech rate as well as the fundamental frequency of the target voice, in a moderate way. This result, typical of the shadowing task, can be interpreted as a latent phenomenon of phonetic convergence, which occurs even in non-interactive communication situations (Delvaux and Soquet, 2007). When the instruction is explicitly to imitate the target voice, the convergence is more marked (the average convergence rate is close to 50% in imitation for the control subjects, both for the rate of speech and for the pitch). On the other hand, convergence is close to zero, or even turns into divergence (speech rate), during the voluntary inhibition of imitation task, thus confirming the ability of the participants to actively adapt their speech productions in order to follow a given instruction. Comparative analysis of convergence rates as a function of conditions (shown in Fig.2) suggests that speech rate is "easier" for participants to manipulate at will than the mean register of f0.

Overall, this study confirms previous results obtained on a population of older French-speaking subjects via a phonetic learning paradigm (Delvaux et al., 2015). The two experimental paradigms are different, as well as the subjects who participated in the two studies, but the finding at the end is similar: even elderly speakers-listeners are able to adapt in real time their speech productions in order to adapt to the internal and external constraints that frame the communication situation. Theoretical implications are numerous, notably for the study of the effects of aging on the motor control of speech (see e.g. Fougeron et al., 2021).

The study presented here also provides an exploratory approach to the question of the potential preservation of phonetic flexibility resources in people with PD. The exploratory nature is due to the small number of participants studied to date, with additional subjects currently being recruited to complete the numbers. Initial results suggest that PD participants exhibit phonetic flexibility, as they do not behave significantly differently from controls, at least for speech rate. On the other hand, MP subjects were found to be unable to converge to (or diverge from) the fundamental frequency of the target voice when instructed. This result could be interpreted in light of the prosodic insufficiency typically observed in these patients (Skodda et al., 2011). Recall, however, the interaction found with the variable Sex: men, with or without PD, do not converge on the pitch, regardless of the experimental task. It is indeed worth noting that the target voice was a female voice (mean pitch at P100: 209 Hz), which makes the imitation/inhibition task radically different for male participants compared to female participants. The small number of participants does not allow at this stage to generalize the effects highlighted by the statistical analysis on our sample: is the lesser convergence on the pitch an effect of gender, an effect of the disease, or more globally an expression of individual variation? Future analyses on a larger sample should answer this question.

In any case, the phonetic flexibility demonstrated for PD patients with respect to speech rate (whereas deficits related to the temporal properties of speech are typical of Parkinsonian dysarthria, Rusz et al., 2011, Skodda et al., 2011) opens interesting clinical perspectives. The speech therapy management of these patients - unfortunately not very systematic in the typical care pathway - could be based on these resources in order to develop the compensatory strategies most adapted to their motor speech disorders.

In conclusion, the exploratory study presented here confirms previous results which suggest that phonetic flexibility capacities are maintained after 60 years of age. These abilities would also be preserved, at least in part, in people with Parkinson's disease. Thus, the production-perception link would remain efficient, mobilizable in order to respond to the constraints of the communication situation, and this despite the presence of a more or less extensive motor execution disorder in these patients.

5. Acknowledgements

This study has partially benefited from the funding attributed to the Research Credit "UltraAeroSpeech", FRS-FNRS, Belgium. Thanks to Yousra El Omari for her help in data collection.

6. References

[1] COOKE M., KING S., GARNIER M., & AUBANEL V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. Computer Speech & Language 28, 2, 543-571. Doi:10.1016/j.csl.2013.08.003.

[2] DELVAUX V., CANO-CHERVEL J., HUET K., LECLEF C., VERHAEGEN C., HARMEGNIES B. (2015). Capacités d'apprentissage phonétique et vieillissement, dans A.-M. ETIENNE (ed). Evolutions sociales, Innovations et Politiques: nouvelles questions pour la psychologie de la santé, Mardaga.

[3] DELVAUX V., SOQUET A. (2007). The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. Phonetica 64, 145-173.

[4] FOUGERON C., GUITARD-IVENT F., DELVAUX V. (2021). Multi-Dimensional Variation in Adult Speech as a Function of Age. Languages 6(4), 176. Doi : 10.3390/languages6040176.

[5] GARNIER M., HENRICH BERNARDONI N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? Computer Speech & Language. Doi: 10.1016/j.csl.2013.07.005.

[6] GRANDJEAN D., BAENZIGER T. (2019). Chapitre 4. Expression vocale des émotions. Dans : David Sander éd., Traité de psychologie des émotions (pp. 109-155). Paris: Dunod. Doi : 10.3917/dunod.sande.2019.01.0109

[7] HUTTUNEN K., KERÄNEN H., VÄYRYNEN E., PÄÄKKÖNEN R., LEINO T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. Applied ergonomics 42, 348-57. Doi: 10.1016/j.apergo.2010.08.005.

[8] MÉNARD L., PERRIER P., AUBIN J. (2016). Compensation for a lip-tube perturbation in 4-year-olds: Articulatory, acoustic, and perceptual data analyzed in comparison with adults. (2016). Journal of the Acoustical Society of America 139(5), 2514. Doi: 10.1121/1.4945718.

[9] MOON S.J., LINDBLOM B., (1994). Interaction between duration, context, and speaking style in English stressed vowels. J. Acoust. Soc. Am. 96, 40–55.

[10] PARDO J. S., URMANCHE A., WILMAN S., WIENER J., MASON N., FRANCIS K. & WARD M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. Journal of Phonetics 69, 1-11. Doi:10.1016/j.wocn.2018.04.001.

[11] PISANSKI K., FRACCARO P., TIGUE C., O'CONNOR J., RÖDER S., ANDREWS P., FINK B., DEBRUINE L., JONES B., FEINBERG D. (2014). Vocal indicators of body size in men and women: A meta-analysis. Animal Behaviour 95, 89–99. Doi :10.1016/j.anbehav.2014.06.011.

[12] REVIS J., DE LOOZE C., GIOVANNI A. (2013). Vocal flexibility and prosodic strategies in a professional impersonator, Journal of Voice 27, 524.e23- 524.e31. Doi: 10.1016/j.jvoice.2013.01.008

[13] ROUAS J.-L., SHOCHI, T., GUERRY M., RILLIARD A. (2019). Categorisation of spoken social affects in Japanese: human vs. machine. Actes de l'International Congress of Phonetic Sciences ICPhS, Aug 2019, Melbourne, Australia. ffhal-02317743

[14] RUSZ J., CMEJLA R., RUZICKOVA H., RUZICKA E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J. Acoust. Soc. Am. 129, 350–367.

[15] SAMEROFF A. J. (2009). The transactional model of development : how children and contexts shape each other. Washington, District of Columbia : American psychological association.

[16] SKODDA S., VISSER W., SCHLEGEL U. (2011). Gender-Related Patterns of Dysprosody in Parkinson Disease and Correlation Between Speech Variables and Motor Symptoms. Journal of Voice 25(1), 76–82. Doi:10.1016/j.jvoice.2009.07.005

[17] SMILJANIC R., BRADLOW A.R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. Journal of Phonetics 36, 91-113.

[18] VERHAEGEN C., DELVAUX V., HUET K., FAGNIART, S., PICCALUGA M. ET AL. (2020). Capacités d'apprentissage phonétique chez des patients aphasiques francophones : étude de cas. Actes des 33e Journées d'Études sur la Parole (JEP), 2020, Nancy, France, 606-616. ⟨hal-02798593v3⟩

[19] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. New York: Springer, 2009.

[21] F. Lastname1, F. Lastname2, and F. Lastname3, "Title of your INTERSPEECH 2022 publication," in *Proceedings INTERSPEECH 2022 – 23rd Annual Conference of the International Speech Communication Association,* Incheon, Korea, Aug./Sep. 2022, pp. 100–104.