Perceptuo-motor biases in the perceptual organization of the height feature in French vowels

Lucie Ménard, Jean-Luc Schwartz

Article type: Scientific paper

Title: **Perceptuo-motor biases in the perceptual organization of the height feature in French vowels**

Authors: Lucie Ménard

Laboratoire de phonétique

Center for Research on Brain, Language, and Music

Département de linguistique, Université du Québec à Montréal

320, Ste-Catherine Est, Montréal, Quebec H2X 1L7, Canada

menard.lucie@uqam.ca

Jean-Luc Schwartz

GIPSA-lab, Département Parole et Cognition

CNRS - INPG/UJF/Université Stendhal,

Grenoble, France

Field:

Classification: PACS: 43.71. An, 43.71.Es

Running head: Height feature in French vowels

Abstract:

This paper reports on the organization of the perceived vowel space in French. In a previous paper [28], we investigated the implementation of vocal height contrasts along the F1 dimension in French speakers. In this paper, we present results from perceptual identification tests performed by twelve participants who took part in the production experiment reported in the earlier paper. For each subject, stimuli presented in the identification test were synthesized in two different vowel spaces, corresponding to two different vocal tract lengths. The results showed that first, the perceived French vowels belonging to similar height degrees were aligned on stable F1 values, independent of place of articulation and roundedness, as was the case for produced vowels. Second, the produced F1 distances between height degrees correlated with the perceived F1 distances. This suggests that there is a link between perceptual and motor phonemic prototypes in the human brain. The results are discussed using the framework of the Perception for Action Control (PACT) theory, in which speech units are considered to be gestures shaped by perceptual processes.

1. Introduction

The rigorous debate between proponents of the auditory and motor theories of speech perception continues (see e.g. [1] vs. [2]), particularly in light of recent neurophysiologic evidence coming from the literature on perceptuo-motor interactions in the human brain [3, 4, 5, 6, 7, 8] drawing arguments both for and against motor theories. Major arguments about the potential role of motor knowledge on speech perception generally come from experiments assessing how the fact that a listener knows something about coarticulation in speech helps her or him better process speech sounds in order to extract invariant features (e.g. [9, 10, 11, 12, 13]). Another more recent set of arguments concerns the possibility of modulating the results of a speech perception task by a perturbation applied to the speech production system, e.g. by transcranial magnetic stimulation [14, 15] or used-induced motor-plasticity [16].

In this paper, we address another type of phenomenon that likely involves speech perceptuo-motor coupling in the human brain. This concerns the possibility that correlations may exist between the way the repertoire of phonemic units for a given subject is implemented in speech perception and speech production.

1.1 Relationships between the perceptual and motor phonemic repertoires

If there is indeed a link between speech perception and action in the human brain, this should result in relationships between the way phonemic repertoires are shaped in

speech perception and speech production. Currently, direct evidence for such links between perceptual and motor repertoires is scarce. Of course, inter-linguistic differences can lead to variations of both motor and perceptual systems, but this produces an apparent link that is the consequence of only one external cause: the fact that the linguistic environment varies according to the subject's language (e.g. between French and Italian subjects), producing differences in both perception and production (see e.g. [17]).

Thirty years ago, a series of studies attempted to demonstrate that subtle differences in speech production and speech perception within a coherent linguistic environment could lead to intra-subject correlations, suggesting a link between *idiosyncrasies* in the perceptual and motor repertories. Bell-Berti et al. [18], quoted in Galantucci et al. [2], reported differences between subjects in the perception of the [i] versus [I] contrast in American English, which seemed to be linked to differences in the articulatory implementation of this contrast, related more with tongue height in one case and tongue tension in the other case. In the same vein, Fox [19] studied perceptual structures of the vowel space in American English through a multidimensional scaling analysis of estimated auditory distances between vowel pairs. He then attempted to correlate individual differences in the use of estimated perceptual dimensions with characteristics of the vowel space in speech production by the same speakers. He claimed to have found some correlations. However, the perceptuo-motor link was not clear, especially since the statistical analysis may have included some artifacts, given the large number of attempted correlations between variables.

During the same period, studies on selective adaptation also provided evidence for dynamic links between motor and perceptual representations (see e.g. [20, 21, 22]). More recently, a series of studies by Perkell and colleagues showed how perceptual abilities can shape production strategies for vowels [23] or fricatives [24], and can adapt to feedback perturbation [25]. On the other hand, more relevant for the present discussion, Shiller et al. [26] recently showed how a perturbation of the production of the [s]-[ʃ] contrast resulted in after-effects with consequences not only on production but also on perception, leading to a boundary shift in the same direction as the one induced on production (see Nasir and Ostry [27] for similar results with both production and perception after-effects following perturbations of the somato-sensory feedback in vowel production).

Such evidence for a dynamic coupling between perceptual and motor behaviors suggests that there should indeed exist some hard-wired coupling between the way phonemic units are represented in the perceptual and motor systems of speech. In order to derive clear-cut evidence for a link between perceptual and motor repertoires in a single subject, the present paper capitalizes on an idiosyncratic effect that we recently discovered related to the production of height contrasts in French vowels.

1.2. F1-series as a paradigm for investigating perceptuo-motor links

In a previous paper [28] we investigated the organization of the height feature in French vowels along the F1 dimension. Twenty-seven French speakers (children around

4 years old, children around 8 years old, and adults) from two dialect regions (France and Canada) were recorded, to generate significant between-subject variability. Each speaker produced repetitions of the ten sustained, isolated French oral vowels /i y u e ø o ɛ œ ɔ a/. Acoustic analyses showed considerable between-subject variability in the relative positions of the vowels within the vowel space. However, speakers tended to produce vowels along a given height degree with a stable F1 value depending on the speaker, but independent of place of articulation and roundedness. Interestingly, this resulted in clear significant differences in the implementation of height contrasts, with some subjects displaying a strong proximity between high (e.g. [i]) and mid-high (e.g. [e]) vowels, while others displayed a strong proximity between mid-high and mid-low (e.g. [ɛ]) vowels, and still others displayed a strong proximity between mid-low and low (e. g. [a]) vowels. Simulations with an articulatory model of the vocal tract (that we will describe later), the Variable Linear Articulatory Model (VLAM), showed that a stable F1 value was related to a stable tongue height. Hence it appeared that each subject displayed her or his own idiosyncratic choice in the precise way the vowel height features were implemented, with a speaker-specific distribution of F1 values inside the vowel triangle.

An individual speaker's idiosyncrasy leading to within-speaker stable F1 values for a given vowel height degree, along with large between-speaker variability of such F1 values, could provide an experimental paradigm for assessing perceptuo-motor links in vowel processing and categorization. The question raised by our recent experiments concerns the role played by this idiosyncrasy in vowel perception processes: Would a subject display as a listener the same type of F1 series as he or she did as a speaker?

Moreover, if this was the case, would there be a relationship between variability in perception and in production, with a similar series of biases in both perceptual and motor vowel prototypes? Finally, how would vocal tract variability intervene in this process? If there was an "F1-series" effect in perception, the distribution of F1 values according to vowel height would be similar for various vocal tract sizes for a given listener. These questions and hypotheses led to the present experiments in which we investigated the organization of the perceived height in French vowels for various listeners—with synthetic stimuli corresponding to various vocal tract sizes and speaker ages—in relation to the listener's own vowel production as a speaker of that language.

1.3 Objectives

We aimed to establish perceptual regions for the French oral vowels, for French listeners ranging in age from 4 years old to adulthood, using stimuli corresponding to two different vocal tract sizes (child and adult). This method would allow us to present listeners with variable formant and F0 values. First, we aimed to study the distribution of mean F1 values for each vowel category to test for possible F1 stability at a given height. Second, we aimed to compare the distribution of F1 values for each vowel category from one vocal tract size to the other for each subject. Finally, we aimed to assess the relationships between production and perception concerning the organization of height degrees along F1.

2. Method

2.1. Participants

The twelve native speakers of Continental French (the CO corpus) who participated in the production study reported by Ménard et al. [28] also participated in the perceptual tests described here. The participants were grouped into three age groups: 4-year-olds (two females), 8-year-olds (two males and two females) and adults (three males and three females). The average ages of the three groups were: 4 years, 10 months (from 3 years, 10 months to 5 years, 10 months); 8 years, 1 month (from 6 years, 2 months to 9 years, 11 months), and 25 years (from 18 to 39 years). These three groups will be referred to as the 4-year-old group, the 8-year-old group, and the adult group. This grouping into three age categories was of course not very accurate, with for example, a small difference in age between the oldest child in the 4-year-old group and the youngest one in the 8-year-old group. However, this very "rough" age grouping did not have strong impact on the experimental paradigm, as discussed later in this section and in Section 3.

None of the speakers reported any history of auditory or articulatory disability. The screening procedure consisted of (1) a brief conversation with the experimenter and a speech language pathologist, (2) a 20-dB pure-tone screening at 500, 1000, 2000, 4000, and 8000 Hz, and (3) for children, a brief developmental test to detect speech production disabilities [29].

2.2. Stimuli

Each subject participated in an auditory identification test. The stimuli for the perceptual experiment were 5-formant vowels that were synthesized with the Variable Linear Articulatory Model (VLAM) [30, 31], an articulatory-to-acoustic model of the vocal tract based on Maeda's adult model [32]. This model integrates data on the longitudinal growth of the vocal tract [33] and has been used previously in phonetic studies [28, 34, 35, 36]. From an articulatory configuration controlled by seven parameters (lip protrusion, lip height, tongue tip position, tongue body position, tongue dorsum position, jaw height, and larynx height), VLAM generates realistic 5-formant vowels corresponding to vocal tract length and configuration representative of speakers from birth to adulthood.

Each subject was presented with two series of synthesized vowel stimuli, one series generated on a vocal tract simulating an age (and hence a vocal tract length) close to the subject's own age—to assess perceptuo-motor idiosyncrasies in optimal conditions—and the other series generated on a vocal tract from a very different age (and vocal tract length), to assess normalization processes.

2.2.1. Adult stimuli

For the first dataset, the adult stimuli, all the stimuli were generated in VLAM, based on vowels spoken by a simulated 21-year-old with a total vocal tract length of 17.45 cm and an F0 value of 112 Hz. Using VLAM, we generated the maximal vowel space (MVS) [37], as described previously [34, 35]. For a given growth stage, the MVS is defined as the maximal acoustic space, in the F1-F2-F3 dimension, obtained by a uniform

distribution of the entire input space of command parameters. All possible vowels of the world's languages can be situated within that space. The stimuli were based on the 38 monophthong oral vowel prototypes in the UCLA Phonological Segments Inventory Database (UPSID), developed by Maddieson [38] and described in Vallée [39]. The UPSID prototypes were an appropriate sample covering the entire MVS, while ensuring articulatory coherence and a reasonable number of stimuli. First, the four corner vowels /i y u a/ were situated at the limits of the MVS, according to the Dispersion-Focalization Theory (DFT) criteria [40]. The DFT assumes that vowel systems are shaped by both dispersion constraints that increase mean formant distances between vowels, and focalization constraints that increase the trend to have focal vowels in the system, that is, vowels with close adjacent formants (F3 and F4 for /i/, F2 and F3 for /y/, F1 and F2 at their lowest mean position for /u/, and F1 and F2 at their highest mean position for /a/). The remaining 34 acoustic prototypes were distributed according to their mean F1, F2, and F3 values provided by previous studies of these vowels in various languages (for a review, see [39]). Once established, the optimal acoustic F1, F2, and F3 values were related to their underlying articulatory values (command parameters in VLAM) by an inversion procedure exploiting the pseudo-inverse of the Jacobian matrix (see [35] for more details). The values of the fourth and fifth formants and the bandwidth values were calculated with the algorithm found in Badin and Fant [41]. The poles of the transfer function were excited through a 5-formant cascade synthesis system [42], by a pulse train generated by a source according to the Liljencrants-Fant (LF) model [43]. The parameters related to the source (glottal symmetry quotient and open quotient) were equal to 0.8 and 0.7, respectively, and remained unchanged for all growth stages. The resulting signal was

sampled at 22,050 Hz and was 500 ms long. Fundamental frequency values were chosen based on the growth data presented by Beck [44]. A rising-falling F0 contour was extracted from a natural vowel production and used to shape the stimuli's F0 contours. Those stimuli were used in previous auditory categorization experiments conducted with adult subjects [28, 35]. This adult dataset, shown in Table I, was presented to each participant.

[Insert Table I]

2.2.2. Non-adult stimuli

Three other sets of synthesized stimuli were generated for three growth stages: newborn, 4 years old, and 8 years old (corresponding to vocal tract lengths in VLAM of 7.74 cm, 10.85 cm, and 12.11 cm, respectively and Fo values of 450 Hz, 337 Hz, and 260 Hz, respectively). To achieve this, MVS were generated for each age; the 38 prototypical acoustic locations of the UPSID database were superimposed on each MVS; and the corresponding F1, F2, and F3 values were used to synthesize the 38 sound files. As was the case for the adult stimuli dataset described earlier, the underlying articulatory command parameters were retrieved using an inversion procedure. As we demonstrated in previous papers [28, 34, 35], a given location within a child's MVS does not necessarily involve the same articulatory settings for the comparable location within the adult's MVS. To ensure a more exhaustive dataset, acoustic results of the articulatory commands determined for the adult stimuli dataset were also generated, yielding an

additional set of 38 sound files, for a total of 76 synthesized stimuli in each of the three younger vocal tracts (newborn, 4 years old, and 8 years old). Those stimuli are depicted, in the F1 versus F2 space, in Figure 1. The formant and bandwidth values of the newborn vocal tract are summarized in Table II.

2.2.3. Distribution of stimuli between listeners

The two 4-year-old subjects were asked to identify stimuli corresponding to a 4-year-old child (76 stimuli, upper-left panel in Figure 1) and an adult (38 stimuli, lower-right panel in Figure 1). The four 8-year-old subjects were asked to identify stimuli corresponding an 8-year-old child (76 stimuli, upper-right panel in Figure 1) and to an adult (38 stimuli, lower-right panel in Figure 1). The six adult subjects were asked to identify stimuli corresponding to an adult (38 stimuli, lower right panel in Figure 1) and to a newborn (76 stimuli, lower-left panel in Figure 1). That is, all participant groups perceived two different sets of stimuli—one close to their age (4 years old, 8 years old, or adult) and one very different from their age (adult for children, newborn for adults). The grouping of children into 4-year-old and 8-year-old age categories enabled us to limit the number of required stimuli. The fact that the age grouping was not very accurate had limited consequences, since the difference between the subject's age and the VLAM age was less than 2 years for each child, which corresponds to a less than 6% difference in vocal tract size.

All stimuli for the newborn, 4-year-old, 8-year-old, and adult (21-year-old) , versions of VLAM are depicted in the F1 versus F2 space in Figure 1.

2.3. Experimental procedure

Participants had completed the production task (described in [28]) before the perception task. In the current study, each participant was presented with a set of random stimuli. The stimuli were presented binaurally via high-quality headphones. The participant was instructed to identify, by selecting an icon on a screen, the perceived vowel among the ten French oral vowels /i y u e ø o ɛ œ ɔ a/. Those vowels represent phonological contrasts along four height degrees: high (/i y u/), mid-high (/e ø o/), mid-low (/ɛ œ ɔ/), and low (/a/). French also features contrasts along the place of articulation (e.g., front [/i y e ø ɛ œ/] or back [/u o ɔ/]) and roundedness (e.g., front unrounded [/i e ɛ/] or front rounded [/y ø œ/]). Each vowel was represented by a monosyllabic word of the structure [fV(C)] : « fil » ([fil]), « fée » ([fe]), « fer » ([fɛʁ]), « fa » ([fa]), « fut » ([fy]), « feu » ([fø]), « fleur » ([flœʁ]), « fou » ([fu]), « faux » ([fo]), « fort » ([fɔʁ]). The test lasted about 10 minutes and took place in a soundproof room. The stimulus presentation was blocked according to the adult and non-adult dataset. The task was self-paced. The inter-stimulus interval was 1 second. For the 4-year-old participants, words were replaced by images of the words or puppets. During a familiarization phase, the experimenter ensured that the child could associate the correct word with the image. During the test, the child pointed to the answer with his or her finger, and the experimenter selected the corresponding image with the computer mouse. The limitation of test duration to 10

minutes – which imposed a limited dataset with only one identification answer per stimulus point – was necessary for this group of very young subjects.

2.4. Data analysis

2.4.1. Perceptual center of gravity

Values of the stimuli's first two formants (F1 and F2) were first converted to the Bark scale, since this scale models the perceptual distribution of frequencies in the human auditory system based on the formula found in Schroeder et al. [45]: $F_{Bark} = 7*\text{asinh}(F_{Hz} / 650)$. Each stimulus was associated with its perceived label. This resulted in associating to each vowel category a series of stimuli perceived as members of this category. Then, the mean F1 and F2 values of the perceived vowel category (corresponding to the center of the category) were calculated.

This method of summarizing a complete coherent identification area by using just its center of gravity has been used in various speech perception experiments since those of Ainsworth [46] (see e.g. [47]). Ainsworth discussed the possible drawback of using this summary, in terms of possible boundary effects (that he called "windowing") and unequal density of the stimuli inside the acoustic space. However, these effects are unlikely to be problematic here considering that (1) comparisons between the front and back areas of the vowel space are based on rather even distributions in these areas, as displayed in Figure 1; (2) comparisons between various ages are based on rather similar distributions from one age to the other, at least for infant/children sets; and (3)

comparisons between distributions from one subject to another in the adult space are based on a single dataset, which is the same for all subjects. Moreover, possible effects related to density or windowing should only result in producing noise or bias, which would possibly *decrease* the significance of tested correlations, but not increase them. Finally, the perceptual data was successfully tested and presented no significant departure from a Gaussian distribution (see Note 1).

2.4.2. Relative position along F1

In order to quantify the distribution of high, mid-high, mid-low, and low vowels along F1, according to the method used by Ménard et al. [28] and Neagu [48], we compared the distance in the F1 dimension, in Bark, between vowel categories of different heights. High, mid-high, mid-low, and low vowels were respectively associated with 1, 2, 3, and 4 degrees. Then the following calculations were carried out, based on the data in Bark:

- For each participant, mean F1 values for each vowel category were computed (x_j, where j is one of the ten French oral vowels /i y u e ø o ɛ œ ɔ a/).
- We defined $m_1 = (x_i+x_y+x_u)/3$ and $m_4 = x_a$ as the minimal and maximal F1 values for the participant in question.
- A normalized index for each of the mid-high and mid-low vowels was computed by the formula: $y_j = 100*(x_j–m_1)/(m_4–m_1), j \in$ /e ø o ɛ œ ɔ/.

A schematic representation of the method, reproduced from Ménard et al. [28], is shown in Figure 2. This method proved to be suitable in the examination of vowel target alignment in the production dataset [28].

[Insert Figure 2]

2.4.3. Statistical analyses

Statistical analyses were conducted to determine if these normalized indexes for mid-high and mid-low vowels were similar across place/roundedness values (stable within-listener F1-series) and variable across listeners (variable between-listener F1-series). For this aim, an ANOVA was carried out on these normalized indices (dependent variable) with height (mid-high or mid-low) and place/roundedness (front unrounded, front rounded, or back) as the within-subject fixed factors, and subject as the random factor. Interaction effects were further explored by planned comparisons using the Bonferroni correction with the alpha level set to 0.05.

Then, correlation analyses were carried out between normalized F1 values for produced versus perceived vowels, for each of the mid-high and mid-low categories, for all subjects. The production data had been acquired previously, and they were reported in a previous study of the CO corpus [28]. The resulting data (Figure 4 in [28]) are displayed in Figure 3. The perception data were provided by extrinsically-normalized indices for stimuli at the corresponding age groups in VLAM (4 years old, 8 years old, or adult) depending on the subject's age (see Section 2.4.2).

[Insert Figure 3]

3. Results

3.1. Perceived height along the F1 dimension

The relative positions of the perceived French vowels along the F1 dimension (indices computed as described earlier in Figure 2) for each listener are presented in Figure 4, for the stimuli generated in the adult vocal tract. In this graph, the relative positions of mid-high and mid-low vowels along the F1 dimension are represented as a percentage of the speaker's F1 range (distance between high and low vowels). Low *y*-values (in the upper portion of the graph) corresponding to mid-high vowels reflect the small F1 distances between high and mid-high vowels, whereas high *y*-values for mid-low vowels stand for the large F1 distances between high and mid-low vowels. For each speaker, the values for the three mid-high vowels are linked by a solid line, whereas the values for the three mid-low vowels are linked by a dotted line. Long solid or dotted lines denote large within-speaker variation in the F1 values within a given height degree. Speakers are sorted along the *x*-axis in ascending order of their *y*-data points for mid-high vowels.

[Insert Figure 4]

Figure 4 shows great between-speaker variability in the relative position of mid-high vowels, as depicted by the position of the solid lines along the *y*-axis. Indeed, values range from 10 (minimal value) to 58 (maximal value), resulting in a between-speaker variability of 48. The same pattern of between-speaker variability was observed for mid-low vowels. *Y*-values for these data points ranged from a minimal value of 47 to a maximal value of 78 (for a between-speaker variability of 31). At first glance, we notice that no effect of age is found, as revealed by the fact that speakers from all three age groups had both small and large *y*-values. Despite this variability, for most participants (A_5, A_2, 4_1, 8_2, 8_1, A_6, 4_2), vowels belonging to the same height degree tended to be perceived with similar F1 values. This pattern is noticeable, on Figure 4, by the fact that the three mid-high vowels /e ø o/ have y-values close to each other and/or the three mid-low vowels /ɛ œ ɔ/ have y-values close to each other. The within speaker differences in y-values between each of the two vowel classes is less than the between-speaker variability for the corresponding height degree. This pattern is less clear for the other speakers (A_3, 8_3, 8_4, A_1, A_4). Nonetheless, a mixed ANOVA conducted on the *y*-values with *height* and *place/roundedness* as within-subject fixed factors and *subject* as a random factor revealed a significant effect of *height*, $F(1,11) = 86.003$, $p < .05$, with mid-high vowels having higher values than mid-low vowels, as expected. The *subject* factor did not have a significant main effect on the variation of indices. No effect of *place/roundedness* was found, as a main effect or in interaction with *height* or *subject*. However, the interaction between *height* and *subject* was significant, $F(11,19) = 4.061$, $p < .05$, with some speakers having smaller perceived F1 values for a given height degree than others.

[Insert Figure 5]

The relative positions of the perceived vowel categories along F1 for the stimuli generated in the newborn, 4-year-old, and 8-year-old vocal tracts are presented in Figure 5. Similar indices as in Figure 4 are presented in Figure 5. As was the case for the data perceived within the adult vocal tract (Figure 4), the perceived data generated in younger vocal tracts show great between-speaker variability. Indeed, for mid-high vowels /e ø o/ (solid line), values range from 7 (minimal value) to 48 (maximal value), resulting in a between-speaker variability of 41. The same pattern of between-speaker variability was observed for the mid-low vowels /ɛ œ ɔ/ (dashed line). *Y*-values for these data points ranged from a minimal value of 40 to a maximal value of 81 (for a between-speaker variability of 41). For most participants (A_3, A_2, A_4, 8_2, A_5, 4_1, A_6), within-speaker variability for a given height degree (as measured by the height of the solid or the dashed line) was smaller than the range of between-subject variability. This pattern was less clear for the other participants. As was the case for the analysis presented above, a mixed ANOVA conducted on the *y*-values with *height* and *place/roundedness* as within-subject fixed factors and *subject* as a random factor revealed a significant effect of *height*, $F(1,11) = 73.609$, $p < .05$, with mid-high vowels having higher values than mid-low vowels, as expected. No effect of *place/roundedness* was found, as a main effect or in interaction with *height* or *subject*. Only the interaction between *height* and *subject* was significant, $F(11,22) = 3.346$, $p < .05$.

3.2. Correlations between perceived F1 values in the adult vocal tract and the younger vocal tracts

The latter finding confirms that the distances between height degrees *y* are variable across speakers but that most speakers tend to align all perceived vowels of a given height along the same F1 values. This is true for various "ages" and accordingly various sizes of the vocal tract used for generating the stimuli, as revealed by Figures 4 and 5. This result may be related to the fact that the alignment pattern does not operate along the raw F1 dimension, but rather at a higher acoustic-auditory level. To explore this hypothesis, relative perceived F1 values (y-values in Figures 4 and 5) in the adult dataset were plotted against relative perceived F1 values in the younger vocal tracts dataset, in Figure 6. In this graph, each speaker is represented by 6 data points (3 relative perceived F1 values for the three mid-high vowels and 3 relative perceived F1 values for the three mid-low vowels). As can be observed in this graph, the relative partition of the F1 dimension for each mid-high and mid-low vowels at the perceptual level for different vocal tract sizes are highly correlated. A linear regression analysis was performed on the data and revealed an r^2 value of 0.77 ($p < .001$), with a regression line close to the first diagonal. Separate linear regression analyses for the mid-high and the mid-low series also showed a high degree of correlation (for mid-high vowels: $r^2 = .52$; for mid-low vowels: $r^2 = 021$) that was significant ($p < .01$) for both correlations. Finally, individual regressions between normalized F1 values for the two sets of stimuli for each listener are displayed in Table III (a), and are significant for the 6 mid-high and mid-low vowels.

[Insert Figure 6]

[Insert Table III]

Hence, for each listener, a specific categorization rule would be applied, based on the listener's internal F1-series templates. Next, we sought to determine how these templates relate to those evidenced in the participants' own vowel production.

3.3. Correlations between produced and perceived F1 values

The results for the current experiment were very similar to our previous findings at the production level [28], for the same subjects. In order to investigate the relationship between produced and perceived vowel height degrees, a linear regression analysis was carried out on the relative position along the F1 values (extrinsically-normalized index) of the produced versus perceived vowels for each of the mid-high and mid-low categories. In a first step, perceived vowels were considered at the same theoretical age in VLAM as the subject's age, in order to match as much as possible the vocal tract characteristics in production and perception. The correlation plots are presented in Figure 7 for the three mid-high vowels for illustration, and all correlation results are provided in Table III (b). They are significant for the six vowels, with rather high r^2 values. The plots in Figure 7 show that the normalized F1 indices are rather close in terms of production and perception for each speaker, as shown by the fact that the values would be adequately fitted by the first diagonal in each diagram (dotted line). Thus, the considerable between-

subject variability at the perceptual level can be largely accounted for by the produced values.

In a second step, perceived vowels were considered at the other VLAM age for each subject (that is adult for children and newborn for adults). Not surprisingly, considering the high correlation between normalized F1 values for the two sets of stimuli for each listener (Table III[a]), correlations were once again significant for the six vowels, with r^2 values ranging from 0.34 to 0.62. Since such correlations between perceived and produced vowel targets are obtained both for auditory stimuli synthesized on a version of VLAM with the same "age" and with a very different age, the "inaccurate" character of age grouping in the three gross categories, discussed in Section 2.2.3, appears retrospectively as not at all problematic.

[Insert Figure 7]

4. Discussion

The results suggest that perceived vowel height in the French language is organized in a way that reveals idiosyncratic patterns. For most participants, vowel targets belonging to similar height degrees, independent of place of articulation or roundedness, are aligned along similar subject-specific F1 values. This pattern echoes the one we found at the production level for the same cohort of speakers [28], which indicates a link between speech perception and speech production, in relation with the debate about auditory versus motor theories of speech perception.

4.1. The perceptual vowel space shaped by production constraints

The existence of a perceptuo-motor link seems clearly displayed in the present data, which raises the question: Does idiosyncrasy in these data come from a production-bias projected onto perception, or from a perception-bias projected onto production?

4.1.1. No clear perceptual basis for the perceptuo-motor vowel height idiosyncrasy

A perceptual origin of the bias seems rather unlikely. First, the F1 values for a given subject generally do not conform to a criterion of maximal acoustic distance. Indeed, these values seldom correspond to 33 and 67, which would mean that vowels of different height degrees were equally spaced along the F1 dimension. Perceptual contrast is nonetheless sufficient [49].

It could be advocated that the specific F1 values corresponding to each vowel height degree might be related to the production-perception link as demonstrated by Perkell and colleagues ([23], [24]). Indeed, the produced distance between height degrees, in terms of percentage of the speaker's F1 range from high to low vowels, may be driven by the perceived distance between the auditory targets related to height degrees. Hence, for a given subject, two series of produced vowels, along the F1 dimension, may be close to each other because the auditory discrimination acuity of this subject between both F1 sets of F1 values is poor. On the other hand, greater distances in F1 between both series of produced vowels for another subject might originate from good auditory

discrimination acuity. However, it is unlikely that this perceptual bias also drives the alignment of the three vowels belonging to the same height degree on similar F1 values. Furthermore, there is no strong reason to expect such large differences between one subject and another in auditory acuity differences along F1, apart from general hearing trouble or special auditory competence, which should not result in strong variations within the F1 space but rather a global superiority in acuity from one subject to another.

4.1.2. A possible motor basis, the "Maximal Use of Available Contrasts" principle

In our previous study about idiosyncrasies in speech production, we introduced a possible motor explanation, which could explain inter-speaker variability, and result in inter-listener variability if perceptual and motor repertories are coupled in some way. This explanation was based on what we called a "Maximal Use of Available Contrasts" (MUAC) principle, extending the concept of "Maximal Use of Available Features" (MUAF) introduced by Ohala [50]. In the course of speech acquisition, once an articulatory control is acquired (such as variation in height degrees for front vowels), it would be easier to integrate new controls (such as variation in place of articulation and roundedness) into already existing ones, namely, to use similar F1 values for various vowels of similar height. Since tongue height can be related to stable F1 values across place of articulation and roundedness [28], the tongue-palate distance would provide a good proximal somatosensory invariant parameter related to vowel height degrees. Thus, producing a mid-high vowel, for instance, involves producing the tongue-palate distance

associated with this specific height degree, independent of place of articulation and roundedness.

The present results suggest that the MUAC principle could also shape the speaker-specific organization of the vowel space in perception. Preferred values of F1 or tongue-palate distances for a given vowel height degree, reflecting the MUAC principle, would act as anchors in the organization of the vowel space for both perception and production, and lead to idiosyncrasy related to child-specific developmental pathways [51], and the developmental pathways are possibly established in relation to the ability to produce a sufficient – though not optimal – perceptual contrast in the vowel system of the selected language. This agrees with a number of exemplar models of speech perception (e.g. [52, 53]) in which one's own speech experience provided by the speech production system would feed the speech perception system with exemplar items incorporating a natural link between auditory and articulatory dimensions in a principled way.

In summary, speech perception would be shaped by an idiosyncrasy in speech production, originating from a developmental motor learning scenario leading to MUAC and stabilizing to a certain extent some inter-individual differences in motor control for vowel height. Of course, this kind of perceptual idiosyncrasy, possibly mirroring motor idiosyncrasy, would not impede a listener to process sounds produced by any speaker of the processed language, by applying all means provided by statistical or dynamic processing of continuous speech. The specific setting of prototypes in the phonemic repertoire stays compatible with the categories of the native language, and hence it

enables the individual to process speech by any speaker of the native language. It can be envisioned, however, that speech from some speakers is easier to process than that from others, and that this variability is listener dependent: some "accents" are closer to the listener's own accent, which likely makes that speaker's speech easier to process than that of other speakers who have a greater number of different idiosyncrasies.

To evaluate this, we are conducting further experiments. In one experiment we are studying the idiosyncrasies inside a given family, to examine whether different idiosyncrasies might exist for brothers and sisters, despite their common phonetic environment during speech acquisition. In another experiment we are assessing the joint modifications of the vowel repertoire in production and in perception in post-lingual deaf subjects after cochlear implantation, to determine if implantation results in slight modification of the implementation of vowel height control, and whether perceptual maps evolve in coordination with motor maps.

4.2. Interpretation in the framework of the Perception for Action Control Theory

The present finding adds some evidence to a number of previous studies suggesting that knowledge about speech production could play a part in speech perception (see a discussion in [13]). While this could be conceived as an argument in favor of "pure" gestural theories (e.g. [9, 11]), other arguments about the role of perceptual processing in the definition of speech production targets and motor behaviors

lead us to claim that the communication units are neither purely perceptual, nor purely motor, but intrinsically perceptuo-motor in nature [54, 55].

In the *Perception for Action Control Theory* (PACT) that we have been developing over the past few years [54, 55, 56]), we endorse a view in which (1), perceptual representations are indeed auditory (or audiovisual) in nature, and not motor, which takes into account the nonlinearities in the articulatory-to-auditory transform [57] shaping speech objects in a perceptually efficient way (e.g. [58]); and (2), perceptual and motor repertoires are, however, jointly shaped in development, inducing implicit articulatory knowledge inside perceptual representations. In PACT, the communication unit through which parity may be achieved is neither a sound nor a gesture but a *perceptually-shaped gesture*, that is, a *perceptuo-motor unit*. In this framework, the possibility that online access to motor resources could enhance perceptual representations remains to be determined (see [59]).

PACT includes the hypothesis of a *co-structuring of perceptual and motor representations* in the course of speech development*, enabling auditory categorization mechanisms to take motor information into account*. While it seems clear that auditory categorization processes exist and develop earlier than speech production knowledge in ontogeny [60, 61], the further development of motor procedural knowledge would, in PACT, *modify* and possibly *enhance* the perceptual repertoire. It is also at this level that subtle online transient modifications of motor or auditory repertories could take place in various kinds of *adaptation paradigms* or *motor resonance phenomena*.

The present work is relevant in this framework. It shows rather strikingly that the motor repertoire is specified differently from one subject to another, and that these differences in motor implementation are accompanied by differences in the perceptual repertoire. The interpretation in the PACT framework is that a developmentally learnt motor repertoire provides a bias to an auditory categorization module.

The fact that there is a correspondence between normalized F1 values in the two vocal tract sizes for each listener (see Figure 6) suggests that there exists an auditory normalization module at some stage in the perceptual architecture. A classical normalization parameter, according to authors such as Disner [62] or Traunmüller [63] is provided by the distance F1-F0 in Bark. In a previous work on normalization parameters for vowel perception [34], we proposed a triplet—(F1-F0) for height, (F2-F1) for frontedness, and F'2 for rounding (all values in Barks)—and showed that these parameters provide efficient normalization parameters for describing synthesized French vowels. In Table III (c) we show that values of F1-F0 in Bark between the two sets of stimuli are highly correlated within listeners for the 6 tested vowels.

Of course, other kinds of perceptuo-motor links may exist; for example, motor schemas may play a role at a lower level, in the auditory characterization per se, before categorization, but this is beyond the scope of the present paper. In any case, the role of some motor knowledge on perceptual processes seems highly likely in view of the present data.

5. Conclusion

In this study, we have shown that the vowel space perceived by French listeners displays stable F1 values for a given vowel height degree (mid high or mid low), although there are large inter-individual differences. This pattern appears stable for stimuli generated in child or adult vocal tract models. The variations in the distribution of F1 values from one subject to another appear to be correlated from perception to production. This adds a new element to the increasing data that suggest there is a link between speech perception and speech production, which is compatible with the Perception for Action Control Theory (PACT) view that speech communication is mediated by perceptuo-motor units that are gestures shaped by auditory (and more generally multisensory) processes.

NOTE

1. Gaussian tests were performed on the distance to regression curves, grouping all vowel categories, with no significant departure from the normality assumption, suggesting a regular distribution of perceptual and motor data: $chi^2(5) = 6.10$, $p = .30$.

Acknowledgments

This work was supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. Thanks to Marlene Busko for copy-editing the paper.

REFERENCES

[1] R. L. Diehl, A. J. Lotto, L. L. Holt: Speech perception. Annual Review of Psychology **74** (2004) 431-461.

[2] B. Galantucci, C. A. Fowler, M. T. Turvey: The motor theory of speech perception reviewed. Psychonomic Bulletin and Review **13** (2006) 361-377.

[3] G. Rizzolatti, L. Craighero: The mirror-neuron system. Annual Review of Neuroscience, **27** (2004) 169-92.

[4] G. Hickok, D. Poeppel: Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. Cognition **92** (2004) 67-99.

[5] G. Hickok, D. Poeppel: The cortical organization of speech processing. Nature Reviews Neuroscience **8 (**2007) 393-402.

[6] A. Lotto, G. Hickok, L. Holt: Reflections on mirror neurons and speech perception. Trends in Cognitive Sciences **13** (2009) 110–114.

[7] S. K. Scott, C. McGettigan, F. Eisner: A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. Nature reviews Neuroscience, **10** (2009) 295-302.

[8] F. Pulvermuller, L. Fadiga: Active perception: sensorimotor circuits as a cortical basis for language. Nature Reviews Neuroscience **11** (2010) 351-360.

[9] A. M. Liberman, I. G. Mattingly: The motor theory of speech production revised. Cognition **21** (1985) 1-36.

[10] J. S. Perkell, D. H. Klatt: Invariance and variability in speech processes. L. Erlbaum, Hillsdale, New Jersey, 1986.

[11] C. Fowler: An event approach to the study of speech perception from a direct-realistic perspective. J. Phonetics **14** (1986) 3-28.

[12] C. Fowler: Listeners do hear sounds, not tongues. Journal of the Acoustical Society of America **99** (1996) 1730–41.

[13] J. L. Schwartz: Filling the perceptuo-motor gap. In C. Fougeron, B. Kühnert, M. D'Imperio, N. Vallée (Eds.): Laboratory Phonology 10. De Gruyter Mouton, 2010, 759-786.

[14] A. d'Ausilio, F. Pulvermuller, P. Salmas, I. Bufalari, C. Begliomini, L. Fadiga: The motor somatotopy of speech perception. Current Biology **19** (2009) 381-385.

[15] M. Sato, P. Tremblay, V. Gracco: A mediating role of the premotor cortex in phoneme segmentation. Brain and Language **111** (2009) 1-7.

[16] M. Sato, K. Grabski, A. Glenberg, A. Brisebois, A. Basirat, L. Ménard, L. Cattaneo: Articulatory bias in speech perception: evidence from use-induced motor plasticity. Cortex **47** (2011) 1001-3.

[17] B. Yang: A study on the production-perception link of English vowels produced by native and non-native speakers. Proc. Eurospeech 2001.

[18] F. Bell-Berti, L. J. Raphael, D. B. Pisoni, J. R. Sawusch: Some relationships between speech production and perception. Phonetica **36** (1979) 373-383.

[19] R. A. Fox: Individual variation in the perception of vowels: implications for a perception-production link. Phonetica **39** (1982) 1-22.

[20] W. E. Cooper, M. R. Lauritsen: Feature processing in the perception and production of speech. Nature **252** (1974) 121–123.

[21] W. E. Cooper: Speech perception and production. Ablex Publishing, Norwood, New Jersey, 1979.

[22] D. G. Jamieson, M. F. Cheesman: The adaptation of produced voice-onset time. Journal of Phonetics **15** (1987) 15–27.

[23] J. S. Perkell, F. H. Guenther, H. Lane, M. L. Matthies, E. S. Stockmann, M. Tiede: The distinctness of speakers' productions of vowel contrasts is related to their discrimination of contrasts. J. Acoust. Soc. Am. **116** (2004) 2338-2344.

[24] J. S. Perkell, M. L. Matthies, M. Tiede, H. Lane, M. Zandipour, N. Marrone: The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. J. Speech Lang. Hear. Res. **47** (2004) 1259-1269.

[25] V. M. Villacorta, J. S. Perkell, F. H. Guenther: Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. J. Acoust. Soc. Am. **122** (2007) 2306-2319.

[26] D. Shiller, M. Sato, V. Gracco, S. Baum: Perceptual recalibration of speech sounds following motor learning. J. Acoust. Soc. Am. **125** (2009) 1103-13.

[27] S. M. Nasir, D. J. Ostry: Auditory plasticity and speech motor learning. PNAS **106** (2009) 20470–20475.

[28] L. Ménard, J.-L Schwartz, J. Aubin: Invariance and variability in the production of the height feature in French vowels. Speech Communication **50** (2008) 14-28.

[29] C. Chevrie-Muller, M. Plaza: Les nouvelles épreuves pour l'examen du langage [New tasks for language evaluation]. Nouvelles Epreuves pour l'Evaluation du Langage (N-EEL), Editions du Centre de Psychologie Appliquée, Paris, France, 2001.

[30] L.-J. Boë, S. Maeda: Modélisation de la croissance du conduit vocal. Espace vocalique des nouveau-nés et des adultes. Conséquences pour l'ontogenese et la phylogenese. In: Journées d'Etudes Linguistiques : « La voyelle dans tous ses états », Nantes, 1997, 98-105.

[31] L.-J. Boë: Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults: consequences for ontogenesis and phylogenesis. Proceedings of the International Congress of Phonetic Sciences 3, San Francisco, 1999, 2501-2504.

[32] S. Maeda: An articulatory model of the tongue based on a statistical analysis. J. Acoust. Soc. Am. **65** (1979) S22.

[33] U. G. Goldstein: An articulatory model for the vocal tracts of growing children. Doctoral dissertation. Massachusetts Institute of Technology, Cambridge, MA, 1980.

[34] L. Ménard, J.-L. Schwartz, L.-J Boë, S. Kandel, N. Vallée: Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. J. Acoust. Soc. Am. **111** (2002) 1892-1905.

[35] L. Ménard, J.-L Schwartz, L.-J. Boë: Role of vocal tract morphology in speech development: perceptual targets and sensori-motor maps for synthesized French vowels from birth to adulthood. J. Speech Lang. Hear. Res. **47** (2004) 1059-1080.

[36] L. Ménard, J. L. Schwartz, L. J. Boë, J. Aubin: Articulatory-acoustic relationships during vocal tract growth for French vowels: analysis of real data and simulations with an articulatory model. Journal of Phonetics **35** (2007) 1-19.

[37] L.-J. Boë, P. Perrier, B. Guérin, J.-L. Schwartz: Maximal Vowel Space. 1st European Conference on Speech Communication and Technology (Eurospeech), 1989, 281-284.

[38] I. Maddieson: Patterns of Sounds. Cambridge University Press, 1984.

[39] N. Vallée: Systèmes vocaliques: de la typologie aux prédictions, Ph.D. Thesis, Université Stendhal, Grenoble, 1994.

[40] J.-L., Schwartz, L.-J. Boë, N. Vallée, C. Abry: The dispersion-focalization theory of vowel systems. Journal of Phonetics, **25** (1997) 255–286.

[41] P. Badin, G. Fant:. Notes on vocal tract computation. Speech Transmission Laboratory – Quarterly Progress Status Report, **2–3** (1984) 53–108.

[42] G. Feng: Vers une synthèse par la méthode des pôles et des zéros. Actes des Journées d'Etude sur la Parole. Paris: Groupe Francophone de la Communication Parlée, 1983, 155-157.

[43] G. Fant, J. Liljencrants, Q. Lin: A four-parameter model of glottal flow. KTH, Speech Transmission Laboratory – Quarterly Progress and Status Report, **4** (1985) 1–13.

[44] J. M. Beck: Organic variation of the vocal apparatus. In W. J. Hardcastle & J. Laver (Eds.), Handbook of phonetic sciences. Oxford: Blackwell., 1996, 256–297.

[45] M. R. Schroeder, B. S. Atal, J. L Hall: Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom, S. Öhman (Eds.): Frontiers of Speech Communication Research. Academic Press, London, 1979, 217-229.

[46] W. A. Ainsworth: Intrinsic and extrinsic factors in vowel judgments. In G. Fant, M. Tatham (Eds.): Auditory Analysis and Perception of Speech. Academic Press, London, 1975, 103-113.

[47] T. M. Nearey: Static, dynamic and relational factors in vowel perception. J. Acoust. Soc. Am. **85** (1989) 2088-2113.

[48] A. Neagu: Analyse articulatoire du signal de parole: caractérisation des syllabes occlusive-voyelle en français. Doctoral dissertation. Signal-Image-Word specialization. Grenoble Institute of Technology (INPG), Grenoble, 1997.

[49] B. Lindblom: Phonetic universals in vowel systems. In J. J. Ohala, J. J. Jaeger, (Eds.): Experimental Phonology. Academic Press, New York, 1986, 13-44.

[50] J. J. Ohala: Moderator's introduction to symposium on phonetic universals in phonological systems and their explanation. Proc. Intern. Congress Phon. Sci. 3, 1979, 181-185.

[51] M. M. Vihman, C. A. Ferguson, M. Elbert: Phonological development from babbling to speech: common tendencies and individual differences. Applied Psycholinguistic **7** (1986) 3-40.

[52] K. Johnson: Speech perception without speaker normalization: an exemplar model. In Johnson & Mullennix (Eds.): Talker Variability in Speech Processing. Academic Press, San Diego, 1997, 145-165.

[53] S. D. Goldinger: Echo of echoes? An episodic theory of lexical access. Psychological Review **105** (1998) 251-279.

[54] J.-L. Schwartz, C. Abry, L.-J. Boë, M. Cathiard: Phonology in a theory of perception for action control. In J. Durand, B. Laks (Eds.): Phonology: from Phonetics to Cognition). Oxford University Press, Oxford, 2002, 255-280.

[55] J.-L. Schwartz, L.-J., Boë, C. Abry: Linking the Dispersion-Focalization Theory (DFT) and the Maximum Utilization of the Available Distinctive Features (MUAF) principle in a Perception for Action Control Theory (PACT). In M. J. Solé, P. Beddor, M. Ohala (Eds.): Experimental Approaches to Phonology. Oxford University Press, Oxford, 2007, 104-124.

[56] J. L. Schwartz, A. Basirat, L. Ménard, M. Sato: The Perception for Action Control Theory (PACT): a perceptuo-motor theory of speech perception. Journal of NeuroLinguistics **25** (2012) 336-354.

[57] K. N. Stevens: On the quantal nature of speech, J. of Phonetics **17** (1989), 3-46.

[58] J. -L Schwartz, P. Escudier: A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. Speech Communication **8** (1989) 235-259.

[59] J.-L. Schwartz, M Sato, L. Fadiga: The common language of speech perception and action: a neurocognitive perspective. Revue Française de Linguistique Appliquée XIII (2008) 9-22.

[60] P. D. Eimas, E. R. Siqueland, P. W. Jusczyk, J. Vigorito: Speech perception in infants. Science **171** (1971) 303-306.

[61] P. W. Jusczyk, J. Bertoncini, R. Bijeljac-Babic, L. J. Kennedy, J. Mehler: The role of attention in speech perception by young infants. Cognitive Dev. **5** (1990) 265-286.

[62] S. F. Disner: Evaluation of vowel normalization procedures. J. Acoust. Soc. Am **67** (1980) 253-261.

[63] H. Traunmüller: Perceptual dimension of openness in vowels. J. Acoust. Soc. Am. **69** (1981) 1465-1475.

FIGURE CAPTIONS:

Figure 1: Stimuli used for the perceptual task, in the F1 and F2 space, in Hertz. The 4-year-old subjects listened to the stimuli depicted in the upper-left panel and in the lower-right panel, the 8-year-old subjects listened to the stimuli depicted in the upper-right panel and in the lower-right panel, and the adult subjects listened to the stimuli presented in the lower panels.

Figure 2: Schematic representation of the metric used to evaluate between-speaker and within-speaker variability in F1 distances between degrees of vowel height. A typical distribution of the perceptual centers for the 10 French oral vowels is displayed. Reproduction from Ménard et al [28].

Figure 3: Mean values of relative position along F1 (as a % of the F1 difference between high vowels and /a/) for the 12 participants in the production task (from Menard et al [28], Figure 4). Data are presented separately for mid-high (/e ø o/, solid line) and mid-low vowels (/ɛ œ ɔ/, dotted line). y_j is calculated as $(x_j–m_1)/(m_4–m_1)*100$, where $m_1 = (x_i+x_y+x_u)/3$, $m_4 = x_a$ and j is one of the six perceived French oral vowels /e ø o ɛ œ ɔ/, for each speaker. For a given height degree and a given speaker, the y-values of the three vowels are linked by a vertical bar. Speakers are sorted along the x-axis in ascending order of their y-data points for mid-high vowels.

Figure 4: Mean values of relative position along F1 (as a % of the F1 difference between high vowels and /a/) for the 12 listeners, for the 38 perceived stimuli generated in the adult vocal tract. Data are presented separately for mid-high (/e ø o/, solid line) and mid-low vowels (/ɛ œ ɔ/, dotted line). y_j is calculated as $(x_j–m_1)/(m_4–m_1)*100$, where $m_1 = (x_i+x_y+x_u)/3$, $m_4 = x_a$ and j is one of the six perceived French oral vowels /e ø o ɛ œ ɔ/, for each speaker. For a given height degree and a given speaker, the y-values of the three vowels are linked by a vertical bar. Speakers are sorted along the x-axis in ascending order of their y-data points for mid-high vowels.

Figure 5: Mean values of relative position along F1 (as a % of the F1 difference between high vowels and /a/) for the 12 listeners, for the 76 perceived stimuli generated in the 4-year-old, 8-year-old and newborn tracts. Data are presented separately for mid-high (/e ø o/, solid line) and mid-low vowels (/ɛ œ ɔ/, dotted line). y_j is calculated as $(x_j–m_1)/(m_4–m_1)*100$, where $m_1 = (x_i+x_y+x_u)/3$, $m_4 = x_a$ and j is one of the six perceived French oral vowels /e ø o ɛ œ ɔ/, for each speaker. For a given height degree and a given speaker, the y-values of the three vowels are linked by a vertical bar. Speakers are sorted along the x-axis in ascending order of their y-data points for mid-high vowels.

Figure 6: Linear regression analysis of the relative positions along F1 of the perceived vowels generated in the adult vocal tract and the perceived vowels generated in the three younger vocal tracts (4-year-old, 8-year-old, and newborn). For each of the twelve speakers, six data points are displayed, corresponding to the three mid-high vowels and

the three mid-low vowels. The correlation is significant and the correlation coefficient (R^2) is 0.77 ($p < 0.001$).

Figure 7: Correlation plots for each of the three mid-high vowels /e ø o/ between the relative positions in F1 in the production dataset and the relative positions in F1 in the perception dataset (listener's age dataset). The solid line corresponds to the regression line, the dashed line corresponds to the line for which $x=y$.