



HAL
open science

Transfer of Learning: What Does It Tell Us About Speech Production Units?

Tiphaine Caudrelier, Jean-Luc Schwartz, Pascal Perrier, Silvain Gerber,
Amélie Rochet-Capellan

► **To cite this version:**

Tiphaine Caudrelier, Jean-Luc Schwartz, Pascal Perrier, Silvain Gerber, Amélie Rochet-Capellan. Transfer of Learning: What Does It Tell Us About Speech Production Units?. *Journal of Speech, Language, and Hearing Research*, 2018, 61 (7), pp.1613-1625. 10.1044/2018_JSLHR-S-17-0130 . hal-01871932

HAL Id: hal-01871932

<https://hal.science/hal-01871932v1>

Submitted on 24 Mar 2019

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

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

Transfer of learning: What does it tell us about speech production units?

Tiphaine Caudrelier¹, Jean-Luc Schwartz¹, Pascal Perrier¹, Silvain Gerber¹ and Amélie
Rochet-Capellan¹

¹ Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab*, F-38000 Grenoble, France

* Institute of Engineering Univ. Grenoble Alpes

Correspondence concerning this article should be addressed to

Tiphaine Caudrelier, Speech & Cognition Department, GIPSA-lab

11, rue des Mathématiques, GIPSA-lab, F-38000 Grenoble, France

Tiphaine.caudrelier@gipsa-lab.fr

Abstract

Purpose: Words, syllables, and phonemes have each been regarded as basic encoding units of speech production in various psycholinguistic models. The present paper investigates the role of each unit in the interface with speech articulation, using a paradigm from motor control research.

Method: Seventy-six native speakers of French were trained to change their production of /be/ in response to an auditory feedback perturbation (auditory-motor learning). We then assessed the magnitude of learning transfer from /be/ to the syllables in two pseudo-words (/bepe/ and /pebe/) and one real word (/bebe/), as well as the after-effect on the same utterance (/be/) with a between-subjects design. This made it possible to contrast the amplitude of transfer at the levels of the utterance, the syllable and the phoneme, depending on the position in the word. Linear Mixed Models allowed us to study the amplitude as well as the dynamics of the transfer and the after-effect over trials.

Results: Transfer from the training utterance /be/ was observed for all vowels of the testing utterances but was larger to the syllable /be/ than to the syllable /pe/ at word-initial position, and larger to the first syllable than to the second syllable in the utterance.

Conclusions: Our study suggests that words, syllables and phonemes may all contribute to the definition of speech motor commands. In addition, the observation of a serial order effect raises new questions related to the connection between psycholinguistic models and speech motor control approaches.

Keywords: auditory perturbation, speech units, auditory-motor adaptation, transfer of learning, syllable

Introduction

Speaking requires the translation of a given message into a sequence of lexical units, that are then associated with a set of time variant motor commands producing the appropriate sequence of sounds. A key step of this process is the decomposition of the lexical sequence into smaller constituents that map onto the motor system. This decomposition is called *phonological encoding*, and has been a topic of interest and debates for years in psycholinguistics. A key issue in these debates is the nature of the units that interface the linguistic input with speech articulation, and two major models have been strongly influential in this field (Dell, 1986; Levelt, 1999)

Dell's model holds that phonological encoding is performed by a hierarchically organized network connecting, from top to bottom, morphemes, syllables, syllabic constituents, phonemes, and features. In this model, the syllabic structure "Onset-Nucleus-Coda" provides the frame of the phonological encoding. The basic phonological units are phonemes associated to their position in the constituents of the syllables, i.e. the onset, the nucleus or the coda. The association of phonemes and their position within the syllable structure, described in terms of phonological features, are suggested to form the interface with speech articulation.

Levelt's model (Levelt, 1989) which also begins at the level of lexical concepts, attributes as well a key role to the syllable in phonological encoding. However, Levelt contests Dell's proposal of a phonological encoding stage that would correspond to an "empty" syllabic skeleton filled by phonemes at a later stage. Levelt's model rather suggests that the earliest level of phonological encoding is the specification of segments in the form of well-identified phonemes, and that the syllabification process integrates which phonemes are the basic constituents of the syllable. This conceptual framework leads to the hypothesis that for frequently-used syllables, the articulatory gestures underlying the production of the syllable as a whole are stored as motor routines in a so-called "mental syllabary" (see also Levelt &

Transfer of learning reveals speech units

Wheeldon, 1994). Thus for the production of frequently used syllables, chunks corresponding to full syllabic motor schemas form the interface with articulation. For new or rarely used syllables, this interface could possibly consist of phonemes that would be assembled on-line into syllables (Cholin, Levelt, & Schiller, 2006).

Conflicting empirical data contributed to raise a ‘syllable paradox’ (Dell, 1986, p. 294; Levelt & Wheeldon, 1994, p. 241) with regards to the role of the syllable in phonological encoding, either as a generic skeleton filled by sub-syllabic units or as fully specified syllabic chunks. Several methods have been used to investigate this issue in psycholinguistics. A classical method is the analysis of speech error patterns. As a matter of fact, most errors are claimed to be exchanges of phonemes sharing the same position within the syllable structure (Shattuck-Hufnagel & Klatt, 1979), suggesting that both a syllable frame level and a phoneme content level operate in conjunction during phonological encoding (Nooteboom & Quené, 2015; Shattuck-Hufnagel, 1992). Processing time in speech production studies has also provided evidence for both the role of syllabic structure and syllabic content. Thus, the repetition speed of two pseudo-words is higher when they share the same syllabic structure in initial position (Sevold, Dell, & Cole, 1995), and masked-priming studies have highlighted the influence of the generic syllabic structure independently of its phonemic content (Ferrand & Segui, 1998; Ferrand, Segui, & Grainger, 1996) as well as the role of the phonemic content in phonological encoding (Manoiloff, Segui, & Hallé, 2016). Besides, latency effects related to the frequency of linguistic units has also been instrumental in the debate related to the nature of speech units. The fact that high-frequency syllables correspond to lower production latencies (Cholin et al., 2006) has been interpreted as evidence for the existence of stored articulatory routines for such high-frequency syllables.

Beyond the ‘syllable paradox’ and debates about syllables vs. phonemes as basic units in phonological encoding, other studies in speech production have shed light on the possible

Transfer of learning reveals speech units

role of larger units such as words. As a matter of fact, it appears that the sound changes affecting phonemes embedded in a word occur at different rates according to the frequency or the grammatical status of the embedding word (Bybee, 2002). Thus, in American English, the deletion of the final stop consonants /t/, /d/ occurs in verbs past participles at a faster rate when the verb is frequently used than when it is not, and it occurs at a slower rate in verbs past participles than in contracted negations, which are among the most frequently used grammatical expressions. These observations were interpreted by Bybee (2002) as evidence supporting the hypothesis that in speech production highly frequent words are not represented as a set of phonemes, but rather as “exemplar clusters of their own” (p. 287). This hypothesis is consistent with those among the exemplar-based theories of speech that consider words as basic phonological units stored in the brain in the form of exemplars, i.e. entities perceived or produced in a given situation, rather than generic prototypes (Välilmaa-Blum, 2009). In this case, word-exemplars would be at the interface between phonology and articulation.

In summary, psycholinguistic studies and speech production data, based on the analysis of speech error patterns, repetition and priming paradigms, as well as frequency and grammatical effects, led to views that might appear contradictory about the nature of the basic phonological units that interface with speech articulation. Two possible interpretations are that various units could contribute in parallel to the definition of speech motor commands, or that, depending on the method used, studies in psycholinguistics could actually look into representations that are not at the same level in the process of phonological encoding during speech production (see Cholin et al., 2006). Thus, speech errors, in particular phoneme exchanges, may give information about units involved in serial order control, but not necessarily about the units ultimately tied to articulation. Priming studies could as well access to levels prior to the interface with speech motor commands.

Transfer of learning reveals speech units

Yet understanding the nature of speech unit(s) enabling the translation of a sequence of lexical units into motor commands has key implications in the clinical field: it would help improving the treatment of various developmental speech disorders and in particular apraxia of speech (AOS). Treatment parameters usually include the selected linguistic unit for training (e.g., phonemes, words or whole speech), along with target selection (e.g., specific phonemes or words to practice pronouncing) to achieve the greatest therapeutic benefits (Crosbie, Holm, & Dodd, 2005). For example, training to articulate an isolated phoneme has often been used by speech therapists in the treatment of severe AOS. However, recent findings suggest that practicing the production of a single consonant does not lead to improvements in the production of this consonant embedded in various syllables produced by those patients (Aichert & Ziegler, 2013). In other words, this training does not transfer to other speech conditions. Choosing the syllable as the target unit seems more efficient since transfer of learning was observed from CV syllables to words containing the specific training syllables.

The notion of transfer of learning is not only at the heart of rehabilitation methods for which recovery observed in speech therapists' offices has to be associated with recovery in the daily life of patients. Transfer of sensorimotor learning from a speech unit to the same unit in a different word or to other units is also an experimental paradigm that could be used with typical speakers to directly address the nature of the representations involved in speech production. Transfer may occur when a common mental representation (e.g., the representation of a specific phoneme or syllable) is used to produce both the learning utterance and the transfer utterance. Hence, according to Houde and Jordan (1998), transfer of sensorimotor learning from a given vowel (e.g. /ɛ/) in a CVC context (e.g. 'pep') to the same vowel in another context (e.g. 'set' or 'forget') is a way to directly assess "the putative hierarchical, segmental control of speech production" (Houde & Jordan, 1998, p.1214). In their seminal study, Houde & Jordan (1998) induced sensorimotor learning in typical speakers using an auditory feedback perturbation

Transfer of learning reveals speech units

paradigm. Subjects were asked to repeat CVC (Consonant Vowel Consonant) words (e.g., “pep”) while their auditory feedback was altered in real time so that the vowel /ε/ sounded more like /a/ (e.g., changing “pep” to “pap”). After many repetitions, a large majority of subjects learned to partially compensate for the perturbation, by pronouncing something closer to “pip”, which suggests a sensorimotor recalibration designated as “*adaptation*”. After the perturbation is stopped, a change in the production of the training utterance can persist. This change, referred to as the “*after-effect*”, demonstrates that the procedure triggers learning. The auditory-motor learning paradigm has been extensively used to study speech motor control (e.g. Max & Maffett, 2015; Purcell & Munhall, 2006). Houde and Jordan also showed a partial transfer of auditory learning from /ε/ in a set of training words to the same vowel in other CVC words suggesting that the vowel could be a speech production unit (Houde & Jordan, 1998, 2002). Similar observations were made later both in English and in Mandarin (Cai, Ghosh, Guenther, & Perkell, 2010; Villacorta, Perkell, & Guenther, 2007).

Transfer hence appears as an experimental tool likely to directly address the relationship between the phonological units involved in phonological encoding and the associated articulatory commands. However, to our knowledge, previous studies of sensorimotor learning transfer involved a single phonological level, and no study has been designed with the specific intent of contrasting the role of different speech units in the specification of speech motor commands. Therefore, we extended the approach introduced by Houde and Jordan (1998) to contrast transfer effects at three levels: global utterances (larger than syllables), CV syllables (regarded as chunks that would be coded globally), and phonemes (that would be coded individually in a given syllabic CV structure). Our experimental protocol consisted of assessing sensorimotor transfer from a given vowel V within a given high-frequency CV syllable to disyllabic CVCV utterances containing this syllable in the first or second position, together with another syllable composed of the same vowel and a different consonant. This protocol enables

Transfer of learning reveals speech units

to investigate whether the decomposition that is tied to speech articulation is constituted of (1) sub-syllabic units (here phonemes) in a given position in the CV syllabic frame, as suggested by Dell's model (1986); (2) syllables encoded as chunks, in line with Levelt's proposal for high-frequency syllables (Levelt, 1989); or (3) other units larger than syllables (words or utterances for instance), as argued by some exemplar theories.

Methods

Participants

Seventy-six native speakers of French (33 females and 43 males, age range 18-35 years) took part in the experiment. Subjects had no reported speech, language or hearing disorders, and were naïve to the purpose of the experiment. French from metropolitan France was their mother tongue and none of them learnt or was extensively exposed to a second language in early childhood (before 6 year of age). Rather than using geographical origin within metropolitan France as a selection criterion, since accents are variable within regions, we explicitly tested the contrast /e/-/ɛ/ in production, as described in the procedure. Two channels were used to find participants: flyers distribution on the campus of University Grenoble Alpes, France, and an e-mail sent via the diffusion list of RISC ('Relais d'Information sur les Sciences de la Cognition') which is a research unity affiliated to the CNRS ('Centre National de la Recherche Scientifique', French national research institute). Anybody can subscribe to this diffusion list on the website of RISC. As a result, participants were students or young professionals, and all of them had a baccalaureate (French high-school diploma). Socio-Economic Status was recorded but it was not a selection criterion since it has not been shown to influence auditory-motor adaptation or transfer, to our knowledge. Participants were split into four groups according to the utterance used to evaluate the transfer (see below). They were randomly assigned to one of the groups, with the female-to-male ratio held consistent among groups.

Equipment and task

The participants were seated in front of a monitor in a soundproof room and wore headphones with a built-in microphone (Sennheiser HME 26-II-600). Utterances were displayed on a monitor. The speakers were instructed to read them aloud in a natural way, without shouting or whispering. They heard their own voice mixed with speech shaped noise in the headphones.

The first two formants (F1 and F2) in speakers' productions were shifted in real-time (with a non-disruptive 14ms delay (Yates, 1963)) using the Audapter system (Cai, Boucek, Ghosh, Guenther, & Perkell, 2008). This auditory feedback perturbation resulted in shifting the vowel /e/ towards /ɛ/ by increasing F1 by 27% and decreasing F2 by 10%.

Experimental design and hypotheses

Training utterance. The training utterance was the monosyllabic logatome /be/ in all groups (written "bé"). The vowel /e/ was selected mainly because of auditory perturbation constraints: the vowel had to be roughly in the middle of two existing French vowels in the (F1, F2) space, and back vowels were rejected because F1 and F2 are harder to detect for such vowels. The consonant /b/ was chosen to limit consonant-to-vowel coarticulation (Recasens, 1997). Several other linguistic factors were also taken into consideration. First, the syllable /be/ is as frequent at the beginning of French words as it is in the middle or at the end, as verified using the Lexique 2 database (New, Pallier, Brysbaert, & Ferrand, 2004). Second, it has been suggested that the magnitude of the adaptation to auditory perturbations is influenced by the frequency of auditory and articulatory neighbors (i.e. the word resulting from the auditory feedback alteration, here /bɛ/, and the word that has to be produced to compensate for it, here /bi/, respectively, Frank, 2011). Accordingly, the training word /be/

was chosen so that its neighbors /bɛ/ and /bi/ were low frequency, thus limiting their influence on adaptation.

Test utterances. As shown in Table 1, the utterances used to evaluate the transfer of sensorimotor learning (“test utterances” henceforth) differed according to the group and were chosen to contrast the transfer at the level of the phoneme, the CV syllable and the word. In Group 1, it was the same as the training utterance /be/ (this group will be referred to as *be_transfer* group below). The pseudo-words /bepe/ and /pebe/ (written “bépé” and “pébé”) were used in two other groups (*bepe_transfer* group and *pebe_transfer* group, respectively). These pseudo-words allowed us to study the transfer to the same syllable in a different utterance (e.g., from /be/ to /bepe/), and the transfer to the same vowel in another CV syllable (e.g., from /be/ to /pe/), while changing the position of the syllable /be/ and /pe/ within the utterance. /p/ is a bilabial stop consonant that is articulatorily very close to /b/ and differs mainly by VOT (Voice Onset Time). The French word /bebe/ (“bébé” which means “baby”) was chosen to evaluate the influence of lexical status (real word vs. pseudo-word) on the transfer in a last group referred to as *bebe_transfer* group. As a matter of fact, lexical status has been shown to influence the magnitude of auditory-motor adaptation (Bourguignon, Baum, & Shiller, 2014). Besides, /pe/ was not selected as a testing utterance since it would not provide any additional information about the research question.

Predictions. The experimental design makes it possible to contrast the transfer magnitude at different levels: same training utterance, same syllable in a different utterance, same phoneme in a different CV syllable. The effect of the position in the utterance together with the lexical status can also be tested.

Our main predictions are that if the interface between phonological encoding and speech articulation is: (1) at the level of phonemes (or any unit smaller than syllables, e.g., phonetic features), then a transfer from the syllable /be/ to the syllable /pe/ should occur; (2)

Transfer of learning reveals speech units

at the level of syllable chunks, there will be more transfer to the syllable /be/ than to the syllable /pe/; (3) at the level of words/utterances, the retention of learning in the utterance /be/ will be greater in magnitude than the transfer to any syllable in the other testing utterances (/bepe/, /pebe/ and /bebe/). Additionally, we have two complementary questions. The position of the syllable within a larger utterance may have an effect on transfer, though we have no specific assumption about the precise nature of this effect. Finally in line with the observations that the perceptual bias towards real-words increases tolerance to auditory perturbation (Bourguignon et al., 2014; Ganong, 1980), we expect transfer to be stronger in the real-word /bebe/ than in the pseudo-words /bepe/ and /pebe/.

Experimental procedure

The experimental procedure is described in Figure 1. In a pre-test (blocks 1 to 3), speakers pronounced six minimal pairs (e.g., real French words “épée” (“sword”) and “épais” (“thick”), pronounced /epe/ and /epɛ/) to explore the contrast /e/-/ɛ/ in each speaker. This was done to be able to control whether it had any impact on the results of the study, since the strength of this contrast is quite variable among French speakers (but so far, no study has shown any impact of a phonemic contrast over the adaptation or the transfer of auditory-motor learning, to our knowledge).

The experiment then consisted of 14 blocks of 20 trials each. At the end of each block, the participants had to press a key to continue the experiment. In the baseline phase (blocks 1 to 3), all groups produced the training utterance (/be/), 3 different utterances, including the training utterance (/be/) and their testing utterance with unaltered feedback. This baseline allows us to measure speakers’ production before learning, providing a reference against which we could assess transfer and after-effect magnitudes. During the training phase, the auditory feedback alteration was gradually set up (block 4) and then held at its maximal magnitude (blocks 5 to 12) while all groups were repeating /be/. The perturbation was then

Transfer of learning reveals speech units

stopped and speakers had to produce the test utterance assigned to their group (transfer phase, block 13). Finally all groups pronounced /be/ again to test for the after-effect (i.e. the trace of the auditory-motor learning and its decline, block 14). It should be noted that in be_transfer group, the utterance is the same in both the transfer and the after-effect blocks.

Measurements

F1 and F2 were assessed with Praat software (Boersma, 1993) using Linear Predictive Coding (LPC) computed on sliding window of 25ms with a 1ms time-step. The values were averaged over 30 consecutive points in the stable part of each vowel production (Rochet-Capellan & Ostry, 2011).

The duration of the voiced part of the syllable, as well as the intensity and the pitch in the stable part of each vowel, were also measured using Praat to control for any effect of prosody on adaptation and transfer in relation to syllable position within words or non-words.

All vowel utterances produced in the various experimental phases were characterized by a global index (F2-F1), with F1 and F2 expressed in Barks, a usual auditory measurement of frequency that represents better human perceptual scale than Hertz ((Schroeder, Atal, & Hall, 1979), see equation (1)).

$$fBark = 7 * \operatorname{argsinh}\left(\frac{fHertz}{650}\right) \quad (1)$$

This (F2-F1) index characterizes the position of a vowel on the open-close dimension for front vowels (Fant, 1960). Then, the main measure chosen to assess transfer magnitude was the percentage of (F2-F1) change compared to the corresponding (F2-F1) value measured in the baseline phase for the same utterance in the same speaker. For instance, the first /e/ of each repetition of /bebe/ in the transfer phase was expressed relative to the average first /e/ of /bebe/ productions in the baseline. This percentage of change will be referred to as (F2-F1) change in the results section.

Statistical analysis – linear mixed models

First, the adaptation in each speaker was estimated by comparing (F2-F1) in the last 40 trials of the training phase to the baseline using a one-tailed paired t-test. This allowed us to check that adaptation occurred in a given speaker. A proportion equality test (using the `prop.test` function of R software) was run to test whether the proportion of adapted speakers was equivalent in each group. The analyses carried out using R software (version 3.2.0 - R Development Core Team, 2016) to evaluate adaptation, transfer and after-effects, were as follows.

General approach with Linear Mixed Models. Since we wanted to assess the evolution of (F2-F1) over trials from the adaptation to the transfer and after-effects blocks, we built several Linear Mixed Models (LMM) using the `lme` function from the `nlme` package of R (Pinheiro, Bates, DebRoy, & Sarkar, 2007). LMMs enabled us to take the temporal structure of our data into account (Liu, Cripe, & Kim, 2009). In all but one case (specified in 2.6.2), we followed a backward deletion approach (Mundry & Nunn, 2009) to select the best linear mixed model using likelihood-ratio test. In all LMMs, (F2-F1) change was the variable to explain and the speaker was considered as a random factor ¹. Post hoc tests were adjusted for multiple comparisons, with the `glht` function from the `multcomp` package of R (Hothorn, Bretz, & Westfall, 2008). In all LMMs, the trial was considered as a numerical factor (ranging from 0 to 19). This factor is used to evaluate the evolution over the block. Evolution is then characterized by a slope through linear regression, assuming a linear variation of (F2-F1) within the block.

¹ As there were repeated measures over time for a given participant, residuals structure could depend on the trial. To check the absence of autocorrelation in residuals, autocorrelograms were plotted. When necessary, the residual variance-covariance structure was modeled to allow within-participant correlated errors.

Main effects. To evaluate the adaptation effects, the considered factors were the group and the trial. The transfer effects required two separate analyses since the experimental design was not fully balanced (e.g., there is no training with /pe/ nor transfer towards a single /pe/). In a first analysis restricted to bepe_transfer pebe_transfer and bebe_transfer groups, we tested the influence of the word, the syllable and the position. This was done by comparing the two best models derived from syllable*position*trial² vs. word*position* trial using AIC (Akaike Information Criterion) that is suitable for non-nested models. In a second analysis, /be/ in be_transfer group was compared to /be/ produced in first position in the other groups.

In order to correct for multiple comparisons in an effective way, the two best models derived from these analyses were merged by adding them. This merge did not affect the regression coefficients of the two models (which were independent). Then we performed post hoc tests.

In the after-effect block, the considered factors were the group and the trial. A linear mixed model was built to test in each group whether the production of /be/ in the after-effect block followed the same trend as the production of the syllable /be/ in the transfer block.

Additional analyses. First, it is important to note that all the analyses performed on (F2-F1) were also performed on (F1-F0) (F1 and F0 also expressed in Bark), another possible index of the open-close dimension, possibly more closely related to mandible aperture since (F2-F1) also involves tongue position (Traunmüller, 1981). However, the analyses with (F1-F0) provided essentially the same results and will hence not be reported in the “Results” section.

In addition, prosodic cues, namely vowel duration, pitch and intensity, were analyzed to assess their potential role in the transfer phase, using three separate LMMs. The considered

² Syllable*position*trial represents the models that include the three mentioned factors, and the simple and double interactions between them.

Transfer of learning reveals speech units

fixed factors were the test utterance and the syllable position. Then, possible correlations across participants between the inter-syllabic difference (first vs. second syllable of each transfer word) in these prosodic cues and the inter-syllabic difference in the transfer magnitude were investigated using Spearman's rank correlation.

Moreover, correlations across speakers between their /e/-/ɛ/ contrast evaluated from the pre-test phase and the magnitude of their adaptation were assessed using Spearman's rank correlation. Contrast in the pre-test was obtained by measuring the formant values for /e/ and /ɛ/ in the six uttered minimal pairs and computing the Euclidian distance between mean values for /e/ and /ɛ/ in the (F1, F2) space. Adaptation was expressed here as the Euclidian distance between the mean values for /e/ in the baseline and at the end of the training phase.

Results

The evolution of (F2-F1) by group over the last block of the training (adaptation), the transfer and the after-effect phases is shown in Figure 2. Figure 3 provides the parameters of the linear mixed model in the transfer block.

Adaptation

We were expecting participants to compensate for the perturbation of their auditory feedback by increasing (F2-F1) in /be/ over the course of the training phase. Since this study is focused on transfer of adaptation, only participants who showed a significant adaptation over 3% were selected in the analysis. Fifty-six speakers satisfied this constraint (~74%), of which 12 in be_transfer group, 15 in bepe_transfer group, 15 in pebe_transfer group and 14 in bebe_transfer group. The proportion of adapted speakers was statistically equivalent in each group ($\text{chisq}(3)=1.25, p=0.74$).

Transfer of learning reveals speech units

At the end of the training phase, the average adaptation was 7.8% (mean) \pm 0.8% (sd). This adaptation was homogeneous among groups and stabilized (Figure 2, left) since the obtained LMM shows no significant effect of the group and no significant change over trials.

Transfer

Figure 2 shows the following trends in the transfer phase. First, some transfer occurred in all groups as evidenced by error bars being all well above the 0 line. The transfer to the monosyllable /be/ (be_transfer group, black markers) is much larger than the transfer occurred in all other groups. It results in a gentle negative slope from the adaptation phase (first dotted line on the top of the figure). In the other groups, the transfer seems higher in the syllable /be/ in initial position (e.g., in /bepe/ or /bebe/) than in /be/ in second position (in /pebe/ or /bebe/) or in /pe/ (in /bepe/ or /pebe/). It is displayed on figure 2 by a lower dotted line starting more or less from the same position as the one in be_transfer group, but with a steeper negative slope. The smallest transfer is observed in the syllable /pe/ in first position (in 'pebe'). This trend is suggested by the lowest dotted line on the figure.

The best model fitting transfer data on disyllabic words actually shows an effect of the produced syllable (/be/ vs. /pe/), of its position in the word (1st vs. 2nd) and of their interaction, on both the intercept and the decay slope. It also highlights the influence of the testing word being identical to the training word (largest transfer in monosyllabic /be/).

Post hoc comparisons reveal that: (1) regarding the intercept (Figure 3A), the transfer in /be/ was significantly higher than in /pe/ when produced in the first position in the word (6.6% \pm 0.9 vs. 3.3% \pm 1.2; $z=4.33$; $p<0.001$); (2) the transfer in /be/ was also significantly higher in 1st position than in 2nd position in the word (6.6% \pm 0.9 vs. 4.3% \pm 0.9; $z=3.72$; $p=0.002$); (3) the decay (Figure 3B) was significantly faster (larger negative slope) in /be/ than in /pe/ when produced in the first position (-0.18% \pm 0.04 vs. -0.07% \pm 0.04; $z=-2.793$; $p=0.045$); (4) the decrease was significantly slower in the test word /be/ than in the first

Transfer of learning reveals speech units

syllable of both disyllabic words /bebe/ and /bepe/ ($-0.06\% \pm 0.07$ vs. $-0.18\% \pm 0.04$; $z = -2.778$; $p = 0.047$). No other significant effects were found.

After-effect

Figure 2 shows that the after-effects in be_transfer and pebe_transfer groups seem to be in continuity with the slow decrease of adaptation along the transfer phase for the monosyllabic /be/ (see the top dotted line on transfer and after-effect). The after-effect for the two other groups appears lower with a faster decay.

The best model fitting after-effect data confirms an effect of the group on the evolution over the block. Post hoc comparisons showed no difference between groups in the intercept. However, the decay in bebe_transfer group was significantly faster than in be_transfer group ($-0.13\% \pm 0.05$ vs. $-0.02\% \pm 0.06$; $z = 2.88$; $p = 0.040$). The after-effect in bepe_transfer group seems to be different than in be_transfer and pebe_transfer groups but this difference was not statistically significant neither with regards to the intercept nor to the slope.

The linear mixed model associating transfer and after-effect data by group showed the continuity in the evolution of (F2-F1) change in be_transfer group over the course of these two blocks ($p > 0.05$). In the other groups, a discontinuity was shown between the transfer and the after-effect blocks ($p < 0.005$ in each of these groups), while the decay speed was not different between blocks ($p > 0.05$ in all groups, including be_transfer group).

Prosodic cues and transfer

The second syllable lasted significantly longer than the first syllable in all disyllabic testing utterances ($+30\text{ms}$; $z = 10.1$; $p < 0.001$). The intensity was higher in the first syllable than in the second one (-1.3 ; $z = -5.85$; $p < 0.001$). The pitch was significantly higher in the first than the second syllable, only in 'bepe' ($+0.07$ Barks; $z = 2.72$; $p < 0.01$) and 'pebe' ($+0.17$ Barks;

Transfer of learning reveals speech units

$z=6.68$; $p<0.001$) but not in ‘bebe’ ($+0.04$ Barks; $z=1.40$; $p=0.16$). The observed prosodic pattern was consistent with the combination in French of a primary stress on the last syllable of a phrase, here associated with a longer duration, and an initial stress on the first syllables, associated with a pitch increase (see Jun & Fougeron, 2002, for more details about French prosody patterns). In addition, no correlation was found between the difference in transfer magnitude of the first vs. second syllable and the inter-syllabic difference in any of these cues in the transfer block (duration: $\rho=-0.02$, $p=0.8$; pitch: $\rho=0.03$, $p=0.8$; intensity: $\rho=-0.09$, $p=0.4$).

Contrast /e/-/ɛ/ in production and adaptation

No correlation was found across speakers between their /e/-/ɛ/ contrast in production (as assessed in the pre-test) and the magnitude of their adaptation ($\rho=0.19$; $p=0.18$).

Discussion

Psycholinguistic models of speech production have proposed conflicting views about the units involved in phonological encoding and providing the interface with speech articulation. We argued that the apparent contradictions between empirical data might be due to the variety of paradigms used to address this question and the fact that most methods do not specifically focus on the interface with speech articulation. In line with Houde & Jordan (1998), the present study investigated the role of various linguistic units in the specification of speech motor commands, by studying their influence on the transfer of auditory-motor learning. It was designed to assess whether motor commands are defined (1) at a sub-syllabic level, which could be phonemes in a given syllabic position, in which case transfer would be observed at the level of the vowel (Hypothesis 1); and/or (2) at the level of syllables, represented as chunks, in which case the transfer magnitude should depend on the syllable (Hypothesis 2); and/or (3) at the level of words or more generally utterances, in which case

Transfer of learning reveals speech units

transfer should depend on the word or the utterance (Hypothesis 3). The results tend to show that transfer of learning occurs at the three levels in parallel. In addition, transfer does not seem to depend strongly on the lexical status of the utterance but is highly influenced by the syllable position in a disyllabic utterance. These findings are discussed in the following sections.

What are the main units of speech production in adults?

First, significant transfer occurred in all groups, in all produced vowels regardless of the embedding syllable. Transfer therefore involves a lower level than the syllable. This level may be the phoneme, here the vowel /e/ that constitutes the nucleus of CV syllables /be/ and /pe/. Importantly, it could also be a phonetic feature, such as the position of the tongue when pronouncing the vowel /e/, or any characteristic common to the syllables /be/ and /pe/ (for example, all syllables with a CV structure, or all syllables with a stop consonant). The present study was not designed to disentangle these possibilities since it contrasted the magnitude of transfer neither across several consonantal and vowel features nor across various syllabic structures. To question the role of a syllabic frame as suggested by Dell's model (1986), in relation to speech articulation, transfer from a phoneme in a given syllabic position (an onset consonant for instance) to the same phoneme in another position (e.g. in coda) should be studied. But the important point here is that sub-syllabic components at the lowest level of the hierarchy of units considered in this study do intervene in transfer patterns. This result is in agreement with previous work showing transfer at the phoneme level (e.g. Houde & Jordan, 1998, 2002, see Introduction). Significant transfer has also been displayed from the vowel /ε/ towards other front vowels such as /i/ or /a/ (Houde et Jordan, 1998) but also towards back vowels (Villacorta et al., 2007), suggesting that broader changes could also occur in auditory-motor learning. This is confirmed in a recent study showing that transfer may occur from one

Transfer of learning reveals speech units

vowel to all the adjacent vowels in the phonetic space – though not towards more distant vowels, see Reilly & Pettibone (2017).

At a higher level in the hierarchy, there was an effect of the syllables (/be/ vs. /pe/) in the magnitude of the transfer. Transfer was lower in /pe/ than in /be/ although the consonants /b/ and /p/ are quite similar from an articulatory standpoint, both of them being bilabial stop consonants, differing mainly in VOT (Voice Onset Time). This result suggests that some transfer specifically occurs at the syllable level and therefore that syllable chunks are involved in the definition of motor commands.

Going one step further, some transfer seemed to occur at the level of a higher-level unit, here the word or the utterance. This claim is supported by conjointly analyzing both the transfer effect and the after-effect. The decay of the transfer in the syllable /be/ of /bebe/ was significantly faster than the decay observed in the monosyllabic /be/ in the transfer block. There was, however, no difference at the beginning of the block (as shown by the equal intercepts in both groups), but only in the decay (i.e., the slope). One could argue that the faster decay of transfer seen in /bebe/ compared with /be/ is driven by the fact that 20 repetitions of /bebe/ actually involved 40 productions of /be/, twice as many as the repetitions of the monosyllabic /be/. This may have led to a faster decrease of the effect magnitude over the transfer phase. If this was the case, there should be a difference in the intercept in the after-effect block: after repeating /bebe/ in the transfer phase speakers should show less after-effect on /be/ than after repeating the monosyllabic /be/. Our results show no difference between the groups in the intercept of the after-effect block. Rather, a break was observed between the production of /bebe/ during the transfer phase and the production of the utterance /be/ in the after-effect: the changes in /bebe/ induced by the adaptation phase progressively decrease during the transfer phase but this does not clearly affect the after-effect on the training utterance /be/. Overall, these observations support the existence of units larger than

Transfer of learning reveals speech units

syllables interfacing with speech articulation. It should be noted that these observations were based on the statistical analysis of the evolution over the transfer and after-effect phases.

Previous works that have assessed the transfer of auditory motor learning in an independent block, after the adaptation phase, have focused only on the first trials of the block (Rochet-Capellan, Richer and Ostry, 2012). Our approach shows that the transfer dynamics itself and the transition to the after-effect bring important insights.

The observed pattern of transfer could also be interpreted in terms of gradient of similarity between the training utterance and the testing utterance. In English, the transfer at a vowel level from one CVC utterance to another was observed to be proportional to the acoustic similarity between the vowels in these CVC as measured before learning (Rochet-Capellan, Richer, & Ostry, 2012). However, in our experiment, no correlation was observed between (1) the transfer and (2) the acoustic distance in the F1-F2 space between the vowel in the training utterance and the vowel in the test utterance in the baseline. Other levels of similarity could potentially explain the patterns observed here, such as an interaction of both categorical variables (e.g., linguistic representations) and more continuous ones (e.g., articulatory or acoustic proximity). Further experimental research and also computational models should help in the understanding of the role of these different variables on transfer in speech learning.

Overall, the results suggest that the transfer of sensorimotor learning does not occur at a specific level in terms of the linguistic unit but rather at several levels in parallel. The interactions between these levels are quite complex. Thus there may be several units that co-contribute to the definition of motor commands. This observation in adults is consistent with Vihman and Croft's developmental theory (Vihman & Croft, 2007) according to which several speech production units may develop gradually and therefore coexist in adults (see also Tilsen, 2016). These results can also be related to the increasingly acknowledged view in

speech perception research that various sublexical and lexical levels coexist and collaborate to ensure the adequate processing of speech inputs in the comprehension of both oral and written language (Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2016).

Real word vs. pseudo-word

Our corpus was also designed to test the influence of the lexical status (real word vs. pseudo-word) on the transfer word. The results do not show any difference in the intercept of the magnitude of transfer between the first syllable of /bebe/ and /bepe/ or between the second syllable of /bebe/ and /pebe/. However, the lexical status has been shown to influence the magnitude of auditory-motor adaptation: if the production of a real word is altered to sound closer to a pseudo-word, the magnitude of the adaptation is supposed to be lower than in the opposite situation (Bourguignon et al., 2014). Our data do not show any evidence of this difference in the transfer phenomena we have observed.

The effect of the position of the syllable in the word

The magnitude of the transfer to the syllable /be/ is significantly higher when /be/ is in the first position of the word than in the second position. The large magnitude of this effect was not expected and our study, to our knowledge, is the first one to display an effect of serial order on the magnitude of the transfer of speech auditory-motor learning. Several explanations could account for this effect – referred to as the position effect.

First, the position effect could be prosody-driven. A stressed syllable could be produced in a more precise way than an unstressed syllable. In most participants, the second syllable lasted longer than the first one, which is consistent with French stress patterns (Jun & Fougeron, 2002). However, no correlation across speakers was observed between the stress pattern and the difference of transfer magnitude between the first and the second syllables.

Transfer of learning reveals speech units

Therefore a prosodic interpretation simply based on stress patterns cannot explain the differences in transfer magnitude driven by the syllable position in the word.

Alternatively, the position effect could be the consequence of an initialization effect. Many studies in motor control have emphasized the specific role played by the initial syllable in the word (MacNeilage & Davis, 2000). Movement initiation is generally regarded as more difficult than movement continuation and is perhaps performed by a specific subsystem within the motor system. Thus the first syllable may be produced more carefully than the following ones, with a better goal specification, and therefore preferentially carry the modification triggered by auditory-motor learning, although the mechanism remains unclear. This could then explain the observation that the transfer in /be/ as the first syllable decays faster than /be/ as a monosyllabic word and /pe/ in first position. At the beginning of the transfer block, the initial syllable would be produced as a separate single syllable, requiring a specific initiation mechanism. Gradually, over several repetitions, initiation would become less difficult and the word would be produced as a whole.

The position effect may also be a serial effect that depends on the auditory perturbation. The training was realized on a monosyllabic word. The perturbed syllable was thus the first one (and the only one) of the utterance. In this context, the speakers may have learned to modify more specifically the first syllable of the sequence. This is consistent with the view that phonological encoding is performed serially and incrementally, as supported by several priming experiments in speech production (Meyer, 1990; Roelofs, 1999).

Finally, a lower degree of transfer to the second syllable than to the first one could be driven by online speech control mechanisms based on auditory feedback (Hawco & Jones, 2010; Tourville, Reilly, & Guenther, 2008). While speaking, a prediction of the auditory outcome based on motor commands is believed to be compared to the acoustic realization. If a difference between the prediction and the acoustic realization is noticed for the first syllable

then online control mechanisms would tend to close this gap and hence to decrease the transfer to the next syllable.

To make an informed decision between these possible explanations of the observed position effect, further investigation is required. The paradigm of auditory-motor learning transfer could be applied to training words consisting of several syllables. Ideally, the auditory perturbation would be applied selectively to a syllable in a specific position in the word (e.g., the second one). The transfer to the same syllable in the first vs. second position of another word would then be evaluated. So far, to our knowledge, the auditory-motor learning paradigm has not been implemented in humans with polysyllabic words. Yet a similar paradigm with a pitch perturbation on a specific syllable in a given position of bird vocalizations was previously applied to zebra finches (Hoffmann & Sober, 2014). In humans, selective perturbation in terms of position in the sequence would raise new perspectives in studying how speech units are put into sequence from both phonological encoding and motor control perspectives.

Limitations and perspectives

The present work provides evidence in favor of a convergence of sub-syllabic, syllabic and supra-syllabic factors in auditory-motor learning transfer, which we have indicated in the previous sections in this Discussion to phonological units likely to be involved in the definition of motor commands in speech production. We have already mentioned in these sections some limitations and perspectives for experimental clarification, in relation for instance with the nature of sub-syllabic units or the effect of syllable position in the word.

More broadly, some methodological limitations intrinsically associated to the auditory-motor learning transfer paradigm could have influenced the results. First, it is important to remind that transfer cannot be studied if there is no adaptation. Hence all our data

Transfer of learning reveals speech units

and conclusions are by nature restricted to subjects actually displaying significant adaptation. The adaptation rate observed in this study is 74%, since 56 speakers out of 76 adapted to the perturbation by significantly increasing F2-F1. Although this adaptation rate does not seem very high, it is very similar to what has been observed in previous studies (Cai et al., 2010; Houde & Jordan, 2002). A possible reason why some speakers do not adapt to the perturbation may be related to the fact that they rely more on their somatosensory feedback than on their auditory feedback during speech production (Lametti, Nasir, & Ostry, 2012). Of course, it would remain to assess by other means that the pattern of superposition of phonological units displayed in the present study for adapting subjects can be extended to all subjects even not displaying similar learning transfer processes.

Whether observed transfer patterns can be extended to all speakers in a given language, is related to more general questions about idiosyncrasies, plasticity and evolution of phonological representations in the course of life. Firstly, the present study is limited to the selection of tested speakers and possibly to the limited range of their age or socio-cultural provenance. Even though there is no indication in previous work that education background or socio economic status has an effect on speaker's magnitude of adaptation or transfer, as they are considered to be pure sensorimotor phenomena, further work could address this kind of question. We cannot also discard at this stage the possibility that the selection of appropriate units driving the definition of motor commands differs among speakers, and that our global data display a mixture of different individual behaviors. However, our experimental design does not allow to assess this hypothesis. More importantly, the theory of developmental phonology (Vihman & Croft, 2007), mainly based on the analysis of infants' babbling and their first produced words, suggests an evolution of speech units over the course of development (see also Tilsen, 2016). According to this theory, words would be the initial basic units of speech production in infants. Syllables and phonemes would then emerge from

Transfer of learning reveals speech units

the observation of similarities in the production of the first words (Cristia & Hallé, 2012).

Words, syllables and phonemes would hence coexist as encoding units in adults. This suggests that auditory-motor learning transfer could provide an adequate paradigm to study the dynamics of the development of the phonological network along life. It also provides a natural tool for dealing with plasticity associated with clinical applications.

Clinical outlook

Our results suggest that transfer of learning can occur at the level of several linguistic units (e.g., the word or the utterance, the syllable, and the phoneme) in parallel. While this study concerns artificial learning driven by a perturbation, treatments of apraxia focus on (re)learning to speak (Maas et al., 2008). If transfer of learning works similarly in these two different situations (which needs to be validated), then the results of the present study suggest that several linguistic units may be targeted in rehabilitation programs. A recent study in four patients with AOS was focused on the transfer of learning at different levels in terms of linguistic units (Aichert & Ziegler, 2013). They observed some transfer from a training syllable to the same syllable embedded in disyllabic words, which is consistent with our results. On the contrary, they did not observe any transfer from the production of an isolated phoneme (a consonant) to the same consonant embedded in a syllable, whereas we found transfer from a vowel in a CV syllable to the same vowel in another CV syllable. This may point out the importance of the context in which the targeted unit is embedded during training. The influence of the context on transfer could be further investigated using the paradigm of auditory-motor learning in unimpaired speakers. In addition, the position effect observed in the current study may draw the attention of speech therapists on the possible influence of the position of the targeted syllable in the word when selecting specific training utterances for their patients. Overall, the present study introduces a method to investigate transfer of

Transfer of learning reveals speech units

learning at various linguistic levels in typical speakers. This could prove useful to fine-tune some aspects of the treatments of patients with AOS.

Conclusion

The purpose of the present study was to investigate the role of several speech units in phonological encoding at the interface between the linguistic input and speech articulation. The transfer of sensorimotor learning from a vowel in a specific context to the same vowel in a different context has been regarded as evidence for the existence of a common representation shared between the training and the test utterances, namely the phoneme (Houde & Jordan, 1998). The present experiment extended this result to other speech units (the syllable and the utterance) by contrasting the transfer magnitude at different levels: (1) to the same vowel in a different CV syllable, (2) to the same syllable in a different utterance, and (3) to the same utterance. The results suggest that words or utterances, syllables and phonemes are all part of the interface with speech articulation.

Additionally, the transfer was greater to the first than to the second syllable of the transfer utterance. This position effect provides new perspectives in regards to the serial control and the sequential encoding of speech production.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152).

References

Aichert, I., & Ziegler, W. (2013). Segments and syllables in the treatment of apraxia of speech : An investigation of learning and transfer effects. *Aphasiology*, 27(10), 1180–

1199. <http://doi.org/10.1080/02687038.2013.802285>

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam, 17*, pp 97-110.

<http://doi.org/10.1371/journal.pone.0069107>

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2014). Lexical-perceptual integration influences sensorimotor adaptation in speech. *Frontiers in Human Neuroscience, 2014 8*, Article 208. <http://doi.org/10.3389/fnhum.2014.00208>

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261–290.

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. In *Proceedings of the 8th ISSP* (pp. 65–68).

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America, 128*(4), 2033–2048. <http://doi.org/10.1121/1.3479539>

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 205–235*. <http://doi.org/10.1016/j.cognition.2005.01.009>

Cristia, A., & Hallé, P. (2012). Global and detailed speech representations in early language acquisition. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech Planning and Dynamics* (pp. 11–38). Frankfurt am Main: Peter Lang.

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder : A comparison of two approaches. *International Journal of Language &*

Transfer of learning reveals speech units

Communication Disorders, 40(4), 467–491. <http://doi.org/10.1080/13682820500126049>

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production.

Psychological Review, 93(3), 283–321.

Fant, G. (1960). Acoustic theory of speech production. In *Chapter* (Mouton, Vol. 2, pp. 119–122). Paris: The Hague.

Ferrand, L., & Segui, J. (1998). The syllable's role in speech production: Are syllables chunks, schemas, or both? *Psychonomic Bulletin & Review*, 5(2), 253–258.

<http://doi.org/10.3758/BF03212947>

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, 723(35), 708–723.

<http://doi.org/10.1006/jmla.1996.0037>

Frank, A. F. (2010). *Integrating Linguistic, Motor, and Perceptual Information in Language Production*. Doctoral dissertation, University of Rochester.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.

Hawco, C. S., & Jones, J. A. (2010). Control of vocalization at utterance onset and mid-utterance: different mechanisms for different goals. *Brain Research*, 1276, 131–139.

<http://doi.org/10.1016/j.brainres.2009.04.033.Control>

Hoffmann, L. A., & Sober, S. J. (2014). Vocal generalization depends on gesture identity and sequence. *Journal of Neuroscience*, 34(16), 5564–5574.

<http://doi.org/10.1523/JNEUROSCI.5169-13.2014>

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <http://doi.org/10.1002/bimj.200810425>

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*

Transfer of learning reveals speech units

(New York, N.Y.), 279(1998), 1213–1216. <http://doi.org/10.1126/science.279.5354.1213>

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: compensation and adaptation. *Journal of Speech Language and Hearing Research, 45*(2), 295–310.

<http://doi.org/1092-4388/02/4502-0295>

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation.

Prabus, 14(1), 147–172. <http://doi.org/10.1515/prbs.2002.002>

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience, 32*(27), 9351–9358. <http://doi.org/10.1523/JNEUROSCI.0404-12.2012>

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223–232. [http://doi.org/10.1016/S1364-6613\(99\)01319-4](http://doi.org/10.1016/S1364-6613(99)01319-4)

Liu, C., Cripe, T. P., & Kim, M. (2009). Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. *Molecular Therapy, 18*(9), 1724–1730. <http://doi.org/10.1038/mt.2010.127>

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*(3), 277–298.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science, 288*(5465), 527–531. <http://doi.org/10.1126/science.288.5465.527>

Manoiloff, L., Segui, J., & Hallé, P. (2016). Subliminal repetition primes help detection of phonemes in a picture: Evidence for a phonological level of the priming effects. *The Quarterly Journal of Experimental Psychology, 69*(1), 24–36.

<http://doi.org/10.1080/17470218.2015.1018836>

Max, L., & Maffett, D. G. (2015). Feedback delays eliminate auditory-motor learning in speech production. *Neuroscience Letters*, *591*, 25–29.

<http://doi.org/10.1016/j.neulet.2015.02.012>.Feedback

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*(5), 524–545. [http://doi.org/10.1016/0749-596X\(90\)90050-A](http://doi.org/10.1016/0749-596X(90)90050-A)

Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: turning noise into signal pollution. *The American Naturalist*, *173*(1), 119–123.

<http://doi.org/10.1086/593303>

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: a new French lexical database. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(3), 516–524. <http://doi.org/10.3758/BF03195598>

Nooteboom, S., & Quené, H. (2015). Word onsets and speech errors. Explaining relative frequencies of segmental substitutions. *Journal of Memory and Language*, *78*(C), 33–46.

<http://doi.org/10.1016/j.jml.2014.10.001>

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. <http://doi.org/10.1037/0033-295X.115.2.357>

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18.

<http://doi.org/10.1080/23273798.2015.1081703>

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2007). The nlme Package.

Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of

Transfer of learning reveals speech units

formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(4), 2288–2297. <http://doi.org/10.1121/1.2173514>

R Development Core Team. (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria*, 0, {ISBN} 3-900051-07-0. <http://doi.org/10.1038/sj.hdy.6800737>

Recasens, D. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America*, *102*(1), 544–561.

Reilly, K. J., & Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *Journal of Neurophysiology*, *118*(5), 2925–2934. <http://doi.org/10.1152/jn.00702.2016>

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory-motor transformations in speech. *The Journal of Neuroscience*, *31*(7), 2657–2662. <http://doi.org/10.1523/JNEUROSCI.6020-10.2011>

Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, *107*(6), 1711–1717. <http://doi.org/10.1152/jn.00773.2011>

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, *14*(2), 173–200. <http://doi.org/10.1080/016909699386338>

Schroeder, M. R., Atal, B. S., & Hall, J. L. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. (ed. & pref. . Lindblom & S. E. G. (ed. Öhman pref., & introd.) (Eds.), *Frontiers of Speech Communication Research* (pp. 217–229). New York: Academic. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=mzh&AN=1983016885&lang=f>

r&site=eds-live

Sevold, C. a, Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*.

<http://doi.org/10.1006/jmla.1995.1035>

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering.

Cognition, 42(1–3), 213–259. [http://doi.org/10.1016/0010-0277\(92\)90044-I](http://doi.org/10.1016/0010-0277(92)90044-I)

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18(1), 41–55. [http://doi.org/10.1016/S0022-](http://doi.org/10.1016/S0022-5371(79)90554-1)

[5371\(79\)90554-1](http://doi.org/10.1016/S0022-5371(79)90554-1)

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39, 1429–1443.

<http://doi.org/10.1016/j.neuroimage.2007.09.054>

Trautmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69(5), 1465–1475.

Acoustical Society of America, 69(5), 1465–1475.

Välismaa-Blum, R. (2009). The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? *CogniTextes. Revue de l'Association Française de Linguistique Cognitive*, 2. <http://doi.org/10.4000/cognitextes.211>

<http://doi.org/10.4000/cognitextes.211>

Vihman, M., & Croft, W. (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, 45(4), 683–725. <http://doi.org/10.1515/LING.2007.021>

<http://doi.org/10.1515/LING.2007.021>

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to perturbations of vowel acoustics and its relation to perception. *Journal of Acoustical Society of America*, 122(4), 2306–2319. <http://doi.org/10.1121/1.2773966>

<http://doi.org/10.1121/1.2773966>

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin*, 60(3), 213–232.

Transfer of learning reveals speech units

<http://doi.org/10.1037/h0044155>

Transfer of learning reveals speech units

Tables and figures

Group		be_transfer	bepe_transfer		pebe_transfer		bebe_transfer		
FACTORS	Testing Word	bé (/be/)	bépé (/bepe/)		pébé (/pebe/)		bébé (/bebe/)		
	Utterance	= Training	≠ from training utterance						
	Lexical status	Pseudo-word					Real word		
	Position in the word	1	1	2	1	2	1	2	
	Syllable	/be/ (= Training)		/pe/ (≠ syllable)		/be/ (= Training)			
	Phoneme	/e/ (=Training)							

Table 1: Experimental design and factors tested. Grey areas indicate when the corresponding factor is identical to the training utterance.

Transfer of learning reveals speech units

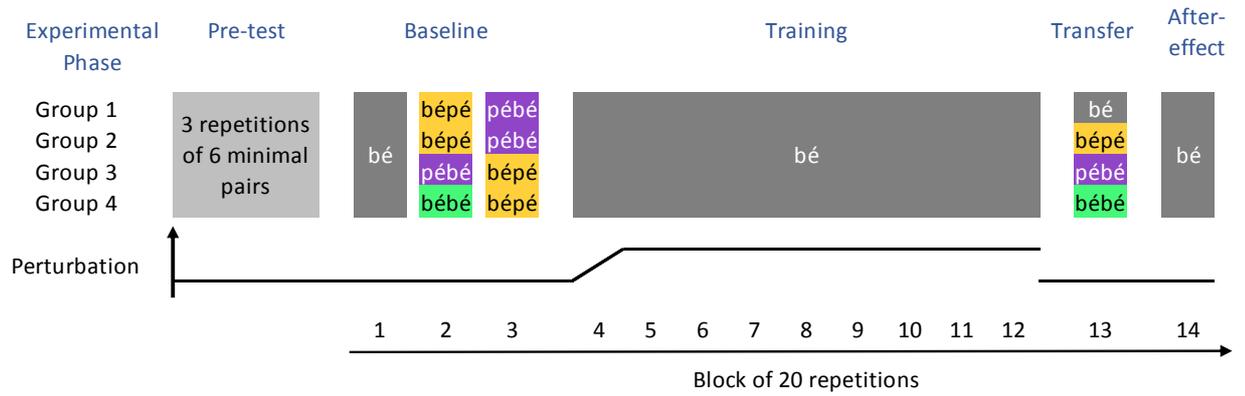


Figure 1: *The experiment consisted of 5 phases, each number corresponding to one block of 20 trials*

Transfer of learning reveals speech units

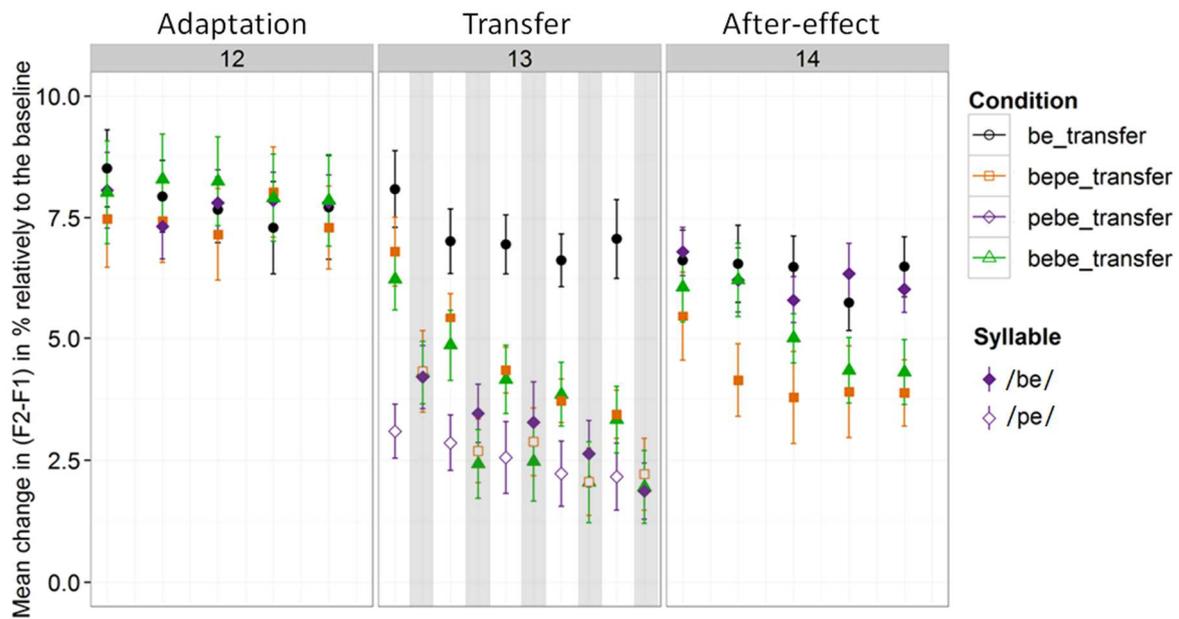


Figure 2: Evolution of (F2-F1) (in % of change in reference to baseline) over the last three blocks of the experiment (corresponding to the end of adaptation, the transfer and the after-effect phases), by group and by syllable. Each point represents the average value for a group and a syllable on 4 trials (with hence 5 points per syllable and per block in a given phase). In the transfer block, the second syllable of the utterance is shifted to the right to separate it from the first syllable, as indicated on the figure (“position 1 vs. 2”). The dotted grey lines highlight the main trends followed by group–syllable–position combinations.

Transfer of learning reveals speech units

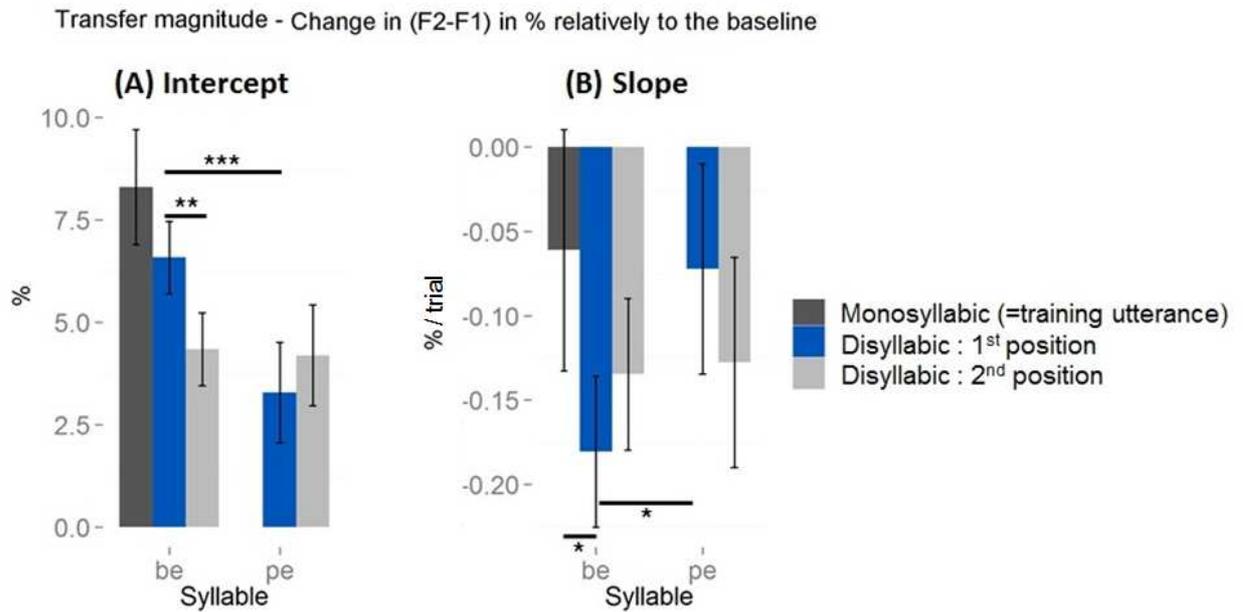


Figure 3: Parameters estimate of the linear mixed model in the transfer phase (A) Intercept; (B) Slope. Results are given by syllable (*/be/* vs. */pe/*), utterance (monosyllabic vs. disyllabic) and by position (1st vs. 2nd syllable in the disyllabic utterances). The bars represent the confidence interval of the estimates. The significance of post hoc comparisons are represented with * (** $p < 0.01$; *** $p < 0.001$; * $p < 0.05$)

Transfer of learning reveals speech units