



**HAL**  
open science

## Cognitive and methodological considerations on the effects of musical expertise on speech segmentation

Clément François, Barbara Tillmann, Daniele Schön

► **To cite this version:**

Clément François, Barbara Tillmann, Daniele Schön. Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Annals of the New York Academy of Sciences*, 2012, The Neurosciences and Music IV Learning and Memory, 1252 (1), pp.108-115. 10.1111/j.1749-6632.2011.06395.x . hal-02062407

**HAL Id: hal-02062407**

**<https://hal.science/hal-02062407>**

Submitted on 8 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.

This unedited manuscript has been submitted for publication in the Annals of the NYAS. This paper has not been copyedited.

**Cognitive and methodological considerations on the effects  
of musical expertise on speech segmentation**

|                               |   |
|-------------------------------|---|
| Journal:                      | <i>Annals of the New York Academy of Sciences</i>   |
| Manuscript ID:                | annals-9999-732   |
| Manuscript Type:              | Conference papers   |
| Date Submitted by the Author: | 31-Oct-2011   |
| Complete List of Authors:     | François, Clément; INCM, CNRS<br>Tillmann, Barbara; Lyon Neuroscience Research Center CRNL, CNRS<br>Schön, Daniele; BDI, INSERM |
| Keywords:                     | statistical learning, language, music   |
|                               |   |

SCHOLARONE™  
Manuscripts

1  
2 **Cognitive and methodological considerations on the effects of musical expertise on speech**  
3  
4 **segmentation**  
5  
6  
7  
8

9 **Clément François<sup>a</sup>, Barbara Tillmann<sup>b</sup>, Daniele Schön<sup>c</sup>**  
10

11 <sup>a</sup>INCM-CNRS UMR6193 & Aix-Marseille University, France  
12

13 <sup>b</sup>CNRS UMR5292, INSERM U1028, Lyon Neuroscience Research Center, Auditory Cognition and  
14 Psychoacoustics Team, Lyon, France & Université de Lyon, France  
15

16 <sup>c</sup>Brain Dynamics Institute, INSERM & Aix-Marseille University, France  
17  
18  
19  
20  
21  
22

23 **Abstract**  
24

25 Both speech and music are constituted by sequences of sound elements that unfold in time, and require  
26 listeners to engage cognitive functions such as sequencing, attention and memory. We recently ran a set  
27 of experiments aiming at testing the effect of musical expertise on a rather high cognitive function:  
28 speech segmentation. Here we will present the main concepts underlying the investigation of speech  
29 segmentation as well as its link to music and musical expertise. Interestingly our results seem to show  
30 that musical training and expertise have effects on brain plasticity that may go beyond primary regions.  
31 Moreover, to facilitate and improve future research in this domain, we will here describe several  
32 delicate methodological precautions that need to be taken into account (e.g., the choice of stimuli,  
33 participants, data analyses). Finally, we will give some possible future directions to better understand  
34 the impact that music may have on speech processing.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## An introduction to statistical learning

Learning the mother tongue or a second language is a rather long process that goes through several dependent phases. Because word boundaries are not systematically flagged by acoustic cues such as stresses or pauses, one important step in language learning is the ability to extract words that unfold in time. Many studies inspired from the seminal work of Saffran and colleagues showed the importance of the statistical structure of the speech stream for an efficient segmentation<sup>1-3</sup>. Indeed, in a speech stream, within-word syllables tend to be associated more often than between-words syllables. The importance of these statistics (called conditional or transitional probabilities) has been shown in adults, infants and neonates<sup>4-7</sup>. The experimental paradigm consists of a familiarisation (learning) phase followed by a test. During the familiarisation phase, participants listen to several minutes of a statistically structured, continuous flow of artificial syllables without any acoustic cues at word boundaries. The test phase depends on the participant population (infants, adults). In the case of adult participants, the test is a two-alternative forced choice procedure (AFC) and participants have to choose, in each trial, between a word that was part of the language and a word built with similar syllables, but that was not part of the language (henceforth partial word). Above-chance performance suggests participants' ability to segment the auditory (speech) stream on the basis of transitional probabilities. This ability has been also demonstrated for non-linguistic auditory sequences: such as sequences of pitches<sup>4,8,9</sup> or of instrumental timbres<sup>10</sup>, thus suggesting that this type of learning is not specific to speech. Finally, this ability has also been shown for sequences of visual stimuli (shapes or movements) and sequences of tactile stimulations<sup>11</sup> suggesting that this type of learning is at work in several modalities<sup>12-14</sup>.

Song is particularly well-suited to have a better understanding of the relations between music and language processing. Indeed, the segmentation of linguistic and non-linguistic inputs has been studied separately (with different tasks and different participants), thus rendering difficult to draw any

1  
2 clear conclusion on the non-specificity of the learning processes for language and music. Recently, we  
3  
4 compared speech segmentation of an artificial sung language (with speech and music combined in the  
5  
6 same signal) to speech segmentation of a spoken language. Speech segmentation was better when sung  
7  
8 than when spoken, possibly due to structural and motivational properties of music<sup>15</sup>. In François and  
9  
10 Schön<sup>16</sup>, participants (nonmusicians) were exposed to the sung speech stream, but then tested  
11  
12 separately on linguistic and musical dimensions of the sung language (both behavioural and EEG  
13  
14 responses were recorded at test). Performance was above chance for the linguistic test, but at chance  
15  
16 level in the musical test. By contrast to this behavioural measure, the analysis of the event-related  
17  
18 potentials (ERPs) revealed in both linguistic and musical tests a similar fronto-central late negative  
19  
20 component that was larger for non familiar (partial words/sequences) than for familiar items  
21  
22 (words/sequences). In agreement with previous findings<sup>3,8,17</sup>, we interpreted this late negative  
23  
24 component as an index for the search of memory traces that have been shaped during learning (i.e., in  
25  
26 the familiarisation phase). In François and Schön<sup>18</sup>, we compared sung speech segmentation in  
27  
28 nonmusicians and professional musicians. While the behavioural results did not show a clear effect of  
29  
30 expertise, ERP data showed a larger late negative component for musicians than for nonmusicians, in  
31  
32 both language and musical tests.  
33  
34  
35  
36  
37  
38  
39

40 Finally, to show that music training was the cause of this difference (and not prior differences of  
41  
42 the participants in the two groups), François et al.<sup>19</sup> conducted a longitudinal study spanning over two  
43  
44 years using a test-training-retest procedure with a pseudo-random assignment of children to two  
45  
46 different artistic training programs. Children followed a training on either music or painting, and were  
47  
48 tested on their ability to extract words from a continuous flow of syllables (the linguistic test). Both  
49  
50 behavioural and electrophysiological measures showed a greater improvement in speech segmentation  
51  
52 across testing sessions in the music group compared to the painting group. Taken together, these  
53  
54 findings point to a benefit of musical expertise and musical training for both speech and music  
55  
56  
57  
58  
59  
60

1  
2 segmentation.  
3  
4  
5  
6

### 7 **Why would speech segmentation benefit from musical expertise?**

8  
9 In 2007, Hickok and Poeppel proposed a dual route model of speech processing<sup>20</sup>, which  
10 includes dorsal and ventral pathways. The dorsal pathway acts as a sensori-motor interface aiming at  
11 mapping the phonologico-acoustic representations of speech sounds to articulatory representations. The  
12 ventral pathway acts as a lexico-conceptual interface aiming at mapping the phonologico-acoustic  
13 representations to lexico-conceptual information. Based on a set of studies combining behavioural,  
14 EEG and fMRI techniques, Rodriguez-Fornells et al.<sup>3</sup> have recently adapted and completed this model  
15 for speech segmentation. They described a large cortical network involved in speech segmentation  
16 comprising the posterior part of the superior temporal gyrus (STG) and the premotor cortex (PMC)  
17 connected via the arcuate fasciculus.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 Within this framework, several non-exclusive hypotheses can explain the differences in speech  
31 segmentation between musicians and nonmusicians. First, musicians might benefit from more efficient  
32 sound encoding at subcortical and cortical levels. Indeed, previous studies have shown functional  
33 differences between musicians' and nonmusicians' encoding of musical and linguistic sounds<sup>21</sup>. In our  
34 experiment<sup>18</sup>, we also found differences due to expertise in rather early auditory components, such as  
35 the N1, the P2 and the mismatch negativity, all possibly generated in the primary auditory cortex and  
36 the planum temporale<sup>22,23</sup>. Thus, early differences in the functioning of the brainstem and the auditory  
37 cortex might explain the musical expertise effect, notably with musical training leading to a  
38 reorganization of auditory neurons along the auditory dorsal pathway, thus facilitating the subsequent  
39 processing steps in the STG and the left inferior frontal gyrus (IFG)<sup>24</sup>.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

54 Second, an alternative explanation could be that musicians may have a more developed and  
55 efficient dorsal pathway than nonmusicians. Its implication for statistical learning has been suggested  
56  
57  
58  
59  
60

1  
2 by a recent study demonstrating that white matter integrity in the vicinity of the left IFG predicted  
3 performance in artificial grammar learning<sup>25</sup>. Similarly, in a speech segmentation experiment  
4 combining fMRI and EEG measures, Cunillera et al.<sup>26</sup> observed a fronto-central late negative  
5 component (at word onsets) that increased over the learning period and that had its generator around  
6 the PMC and the Left IFG. Taken together with the observation of increased grey matter density and  
7 volume in the left IFG for musicians in comparison to nonmusicians<sup>27</sup>, these results suggest that  
8 musicians' advantage in speech segmentation might be related to enhanced involvement of the pre-  
9 motor brain areas in comparison to nonmusicians.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

20  
21 Third, we might hypothesize that musical training improves the connectivity between these two  
22 subsystems. This hypothesis can be integrated in the recent proposition that sound acts as a scaffolding  
23 framework for cognitive sequencing<sup>28</sup>, also supporting how to process and interpret sequential and  
24 temporal information in the environment. It is suggested that sound and speech processing has an  
25 additional unspecified influence on the development of general cognitive sequencing abilities (also in  
26 other modalities). While Conway et al.<sup>28</sup> find support for their hypothesis in the consequences of  
27 auditory deprivation on domain-general sequencing (i.e., impaired implicit learning for visual, non-  
28 linguistic regularities in deaf children), additional support can be found in research investigating the  
29 effect of musical expertise, that is increased training on sound analyses and enhanced sound exposure  
30 should lead to improved sequencing also in non-musical tasks (such as speech segmentation).  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

### 47 **Methodological considerations**

#### 48 (i) The comparison of musicians and nonmusicians

49  
50 When comparing the performance of musicians and nonmusicians, one has to keep in mind  
51 some general difficulties and restrictions to conduct proper experiments testing for the effect of musical  
52 practice. Indeed, most studies comparing musicians and nonmusicians use a cross-sectional approach: a  
53  
54  
55  
56  
57  
58  
59  
60

1  
2 group with 10 years of musical practice is compared to a group without musical practice possibly  
3  
4 comparable in age, handedness, sex and level of education. These studies are very instructive, but  
5  
6 always fall in the criticism of non-controlled pre-existing genetic or other factors explaining the  
7  
8 observed between-group differences. These issues can be ruled out with a longitudinal approach testing  
9  
10 naïve (with respect to music) participants enrolled in a test-training-retest procedure, while still  
11  
12 controlling for several standardized neuropsychological tests and several socio-economic variables.  
13  
14 After the first testing session, this specific procedure requires a pseudo-random assignment of  
15  
16 participants to either a musical group or a control group. This procedure also encounters some  
17  
18 difficulties, like the critical choice of the activity proposed to the control group and the repetition of the  
19  
20 tests which renders the test explicit. This is particularly delicate when interpreting speech segmentation  
21  
22 as an implicit learning process (cf. Section iii).  
23  
24  
25  
26  
27  
28  
29

30 (ii) The choice of the instructions

31  
32 The instructions before the learning phase can become, in our opinion, important because they  
33  
34 determine to which extent learning may take place implicitly or not. For instance, in some studies,  
35  
36 participants are explicitly told about the presentation of an invented “non-sense” language, or even to  
37  
38 look for the words embedded in the speech stream (e.g., Saffran et al.<sup>29</sup>: “figure out where the words  
39  
40 begin and end”). Another study present the isolated words (visually) before the familiarisation phase in  
41  
42 order to maximize learning<sup>17</sup>. By contrast several studies, like ours, do not give any explicit  
43  
44 information before the familiarisation phase: participants are asked to carefully listen to a continuous  
45  
46 stream of syllables for several minutes.  
47  
48  
49  
50  
51  
52  
53

54 (iii) Measuring learning: from explicit to implicit measures

55  
56 Another methodological difficulty refers to the implicit nature of the tests and, of course, the  
57  
58  
59  
60



1  
2 implicit nature of the learning process. Studies investigating speech segmentation mostly refer to  
3  
4 “statistical learning” rather than “implicit learning”<sup>30</sup>, which leads to the absence of testing whether the  
5  
6 learned knowledge is implicit. The domain thus adapts a definition of implicit learning as the incidental  
7  
8 nature of the acquisition process and without the intention to learn (at least for some of the instructions,  
9  
10 see above), rather than extending it to the implicitness of the acquired knowledge<sup>31,32</sup>.

11  
12  
13  
14 The standard behavioural test used in this domain can be criticized for theoretical purposes in  
15  
16 the field of implicit learning. The two-alternative forced choice procedure requires participants to make  
17  
18 an explicit judgement on the two presented items without feedback (“indicate which of the two strings  
19  
20 sounded more like a word from the language you have heard before”). However, the representations of  
21  
22 the segmented items might be weak and might vanish rapidly in time, probably also due to the  
23  
24 interference caused by the presentation of non-familiar items. Thus, there is a need to promote implicit  
25  
26 measures, like ERPs, which do not necessary require an overt behavioural response (e.g., analysis of  
27  
28 ERPs during the familiarisation phase). The priming paradigm could also be adapted to the testing of  
29  
30 speech segmentation as it does not require telling participants about the structure of the material and as  
31  
32 it has already been successfully applied to artificial grammar learning<sup>33</sup>. Similarly, the Serial Reaction  
33  
34 Time procedure can be also used as a promising implicit measure, as introduced by Hunt and Aslin<sup>12</sup>  
35  
36 for the visual modality with 3-element units: Over exposure, response times decreased for elements  
37  
38 within the units compared to response times for elements crossing unit-boundaries.  
39  
40  
41  
42  
43

44  
45 Another relevant point to consider is the use of repeated tests in within-subject designs, such as  
46  
47 for test-training-retest designs or designs comparing different types of languages or studying the  
48  
49 transfer from one language structure to another (see, for example, Omigie & Stewart<sup>34</sup>). Indeed, when  
50  
51 interpreting results and comparing them to other results reported in the literature, one needs to be aware  
52  
53 of the fact that once tested with a typical AFC task, learning of a successive language structure will be  
54  
55 even less implicit, insofar as subjects are then aware that the stream contains words and possibly of the  
56  
57  
58  
59  
60

1  
2 structure of the stream (e.g., trisyllabic words). One possibility to at least blur the cues for potential  
3  
4 strategies gained in the first test phase and then orienting the perception of the second exposure phase  
5  
6 is the use of test items of different lengths (bi- and trisyllabic items), even though only trisyllabic words  
7  
8 are relevant<sup>35</sup>.  
9  
10

11  
12  
13  
14 (iv) The control of the stimuli (units and stream)  
15

16 When preparing an artificial language, great care must be taken about the choice of the  
17  
18 elements, the definition of the units and their chaining in the stream. The first choice is related to the  
19  
20 choice of the phonemes. For instance, a language containing consonants t, d, b, p and vowels i, u, will  
21  
22 be more difficult to segment than a language containing t, s, m, p and a, o, due to the greater phonetic  
23  
24 proximity in the former compared to the latter. Also, when planning to acquire EEG, one may want to  
25  
26 use consonants with a short and similar attack time (e.g., plosives) in order to have clearer ERPs (N1-  
27  
28 P2 complex) to the onset of syllables/words<sup>3</sup>.  
29  
30  
31

32  
33 Another important parameter is the number of words to be used to build the stream (typically  
34  
35 between 4 and 6). Because the stream is typically generated by a pseudo-random concatenation of the  
36  
37 words (no repetition of the same word twice in row), using a very small number of words has a direct  
38  
39 impact of the transitional probabilities at word boundaries. In the extreme case of using two words (eg.  
40  
41 bada-tibu), the transitional probability at word boundaries (da-ti, bu-ba) would be identical to the  
42  
43 transitional probabilities between the two words. This might be particularly relevant, when using partial  
44  
45 words in the test phase (in particular for the 2-AFC test), because, when using a few words, TPs for the  
46  
47 partial words will be very close to TP for words (identical in the case of 2 words). Thus, one may want  
48  
49 to calculate TPs for words AND partial words before running the experiment. Also note, that while  
50  
51 most studies<sup>35</sup> used only trisyllabic words, some studies started to use words of different lengths aiming  
52  
53 to get closer to real-world language learning situations<sup>35</sup>.  
54  
55  
56  
57  
58  
59  
60

1  
2 Another parameter that needs to be taken into account is the ratio between the duration of the  
3  
4 familiarisation phase and the number of words, weighted by the number of syllables building a word.  
5  
6 These choices can vary considerably across the published experiments, going from a 2-minute stream  
7  
8 using 4 bisyllabic words (in babies<sup>1</sup>) up to a 21-minute stream using 6 trisyllabic words (in adults<sup>29</sup>).  
9  
10 One may also note that while behavioural experiments often use rather long and boring familiarisation  
11  
12 durations, EEG analysis seems to show that learning may take place in the very first minutes<sup>26,37</sup>. This  
13  
14 suggests that the 2-AFC might not be a sensitive measure of implicit learning, as already discussed  
15  
16 above (cf. Section iii).  
17  
18  
19

20  
21 Another decision to be taken concerns the definition of the 2-AFC test phase: participants have  
22  
23 to choose between one item that is a word and the other item that is not. For the latter one, three types  
24  
25 are commonly used: non-words for which syllables are arranged in an order that has never occurred in  
26  
27 the familiarisation phase; partial words for which two of the three syllables have appeared in that order  
28  
29 in the familiarisation phase, but the association with the third syllable has never been heard; and partial  
30  
31 words that have been heard in the familiarisation phase, thus containing the boundary between two  
32  
33 words (e.g., da-ti in the 2-words language example given above). Of course, the choice of one of the  
34  
35 three test item types influences the interpretation in terms of learning processes. When using non-  
36  
37 words, participants may simply rely on the detection of a new transition between two syllables, while  
38  
39 the use of partial words discards this possibility and is thus more informative for statistical learning  
40  
41 (see<sup>38</sup> for a more complete discussion on this issue).  
42  
43  
44  
45  
46

47 For the speech material preparation, another delicate point that is rarely addressed in the  
48  
49 literature is the synthesis procedure used to generate the stream. Indeed, because acoustic cues need to  
50  
51 be controlled, a voice synthesizer is used to generate speech (even though cross-splicing techniques  
52  
53 might represent an alternative method). However, speech synthesis is far from being a simple affair and  
54  
55 it is thus important to carefully check the acoustic features of the generated stream. This is even more  
56  
57  
58  
59  
60

1  
2 important when using a sung language, insofar as there might be unexpected interactions between  
3  
4 vowels and pitch, resulting in perceptual accents or lengthening as well as in more or less clearly  
5  
6 pronounced consonants.  
7

8  
9 The importance of a careful selection in the material construction is not restricted to verbal  
10  
11 material only, but also affects non-verbal segmentation tasks. In the case of music, for instance tone  
12  
13 triplets may pop out within the stream (reinforcing or obscuring boundaries). When tone sequences (or  
14  
15 sung sequences) are used, special care must be taken to the choice of the pitches (and pitch classes),  
16  
17 their potential link with musical structures (e.g., tonality), but also the defined interval sizes and  
18  
19 interval directions between adjacent tones (e.g., jumps, reversals) either within units or across unit  
20  
21 boundaries (see also<sup>4</sup>). When musical timbres are used, care must be taken regarding acoustic similarity  
22  
23 (or more generally surface features) for the definition of units and boundaries (see<sup>10</sup>). It is worth  
24  
25 underlining that these indications only concern the choice of the various features in relation to the  
26  
27 definition of units and their boundaries in the stream. Nonetheless, implicit learning can be studied with  
28  
29 materials varying in distance and similarity with real-life materials (e.g., Loui et al., showing artificial  
30  
31 grammar learning with a new musical scale<sup>39</sup>).  
32  
33  
34  
35  
36

37 Regarding the selection of elements and units, some experimental controls have been proposed  
38  
39 (though not systematically applied yet) to ensure that above chance performance is not due to some  
40  
41 preference bias of syllable (or tone) combinations in statistical units or other unrelated influences. 1)  
42  
43 Saffran et al.<sup>4</sup> proposed to define two language systems (L1, L2) that were built in such a way that L1  
44  
45 consists of partial units of L2, and vice versa. This construction allows using the same test phase for  
46  
47 two participant groups (having been exposed to one or the other language), and thus avoiding a  
48  
49 confound between learning and other influences (perceptual biases, preferences, etc...). 2) Reber and  
50  
51 Perruchet<sup>40</sup> suggested to use multiple implementations: all languages have the same statistical  
52  
53 structures, but are instantiated by different syllable attributions (see also<sup>41</sup>).  
54  
55  
56  
57  
58

(v) Data analyses

In several statistical learning studies, t-tests are used to make statistical inferences on whether learning has taken place or not. In particular, group performance is compared to chance level. While the two sample t-test is not appropriate because a “chance” sample has no variance, one should use a one sample t-test, setting up a normal distribution with a mean specified by the null hypothesis (here 0.5). Also, non-parametric tests, such as the one-sample Wilcoxon Signed Rank test, might be more appropriate. Another (non-exclusive) approach is to have either a between-subject design directly comparing results of two or more samples/populations or a within-subject design directly comparing, for instance, learning of the linguistic dimension and the musical dimension, or learning before and after music training (however, see above for comments on repeated-measure designs). Finally, a solution at the individual subject level might be to use binomial tests to estimate the exact probability for each individual. Then, one may run second-order statistics on the p-values, although this might lead to very conservative test (with 36 trials the individual threshold for significance would be 24 correct responses,  $p = .0326$ ).

Another type of analysis consists in taking into account the differences between items. Because each item has average transitional probabilities between syllables (TPs), it is possible to test whether items with high TPs have been learned better than items with low TPs. At this aim, one may either run a non-parametric ANOVA with items as a factor, or contrast directly high versus and low TPs items. These analyses should point out the importance of TPs, by showing poor performance for low TP items and good performance for high TP items<sup>29</sup>.

To better define the role of the TPs in the learning process one could also take into account the distinctiveness between the two items in each trial. This is of specific interest when using partial words that have been heard during the familiarisation phase (i.e.,  $TPs > 0$ ) and that by consequence may also

1  
2 have a sort of pre-lexical representation competing with words. The underlying idea is that trials with a  
3  
4 high TP contrast between words and partial words (i.e., very different TPs) should be easier and thus  
5  
6 show better performance than trials with a low TP contrast (i.e., very close TPs).  
7  
8

9  
10 Finally, another interesting way to model data is, in our view, taking into account time. As we  
11  
12 discussed above, the test is rather explicit and can lead to interference or learning effects along the  
13  
14 testing session (typically 36 trials). Thus, modelling time in the statistical analyses by comparing  
15  
16 performance in the first half of the test to performances in the second half of the test can be very  
17  
18 instructive in this respect. For instance, Rohrmeier et al.<sup>42</sup> used an artificial grammar paradigm with  
19  
20 sequences of tones and showed increasing performance along the testing session for the untrained  
21  
22 group while the trained group showed decreasing performance along the test.  
23  
24  
25  
26  
27

### 28 **Speech segmentation and musical expertise: perspectives and future directions**

29

30  
31 This last section presents possible future directions in the field of musical training and/or  
32  
33 practice-shaped brain plasticity and statistical learning, keeping in mind that the most important point,  
34  
35 in our view, is understanding what aspects of music training (and its consequences on the brain) might  
36  
37 contribute to beneficial effects for speech segmentation and to what extent. Several non-exclusive  
38  
39 directions might be interesting in this respect. For instance, aiming to determine the relevant processes,  
40  
41 one could study the effect of different types of musical trainings, such as rhythmic training versus  
42  
43 pitch/tonal training, by comparing, for example, a group of drummers to a group of singers.  
44  
45  
46

47  
48 Another promising perspective will be investigating the relation between perceptual and  
49  
50 productive musical skills and their interaction with speech segmentation processes. For this aim,  
51  
52 psychophysical measures of perceptual and productive skills could be correlated with statistical  
53  
54 learning performance as well as with EEG or fMRI markers of speech segmentation collected on the  
55  
56 same participants. Furthermore, because musicians' advantage could be in part explained by a better  
57  
58

1  
2 pre-attentive sound processing, other experiments could record both pre-attentive and attentive  
3  
4 subcortical (e.g., frequency-following responses) and cortical auditory responses (e.g., MMN and  
5  
6 P300) and correlate these measures with behavioural performance and psychophysical tests.  
7

8  
9 Another possible direction is to test interactions between musical expertise and different  
10  
11 acoustic and linguistic properties of the speech segmentation input. For instance, it has been shown that  
12  
13 lexical stress (including changes in pitch and timing) and subliminal gaps at word boundaries facilitate  
14  
15 speech segmentation (e.g.,<sup>37,42,43</sup>). Thus, it could be interesting to see to what extent musical practice  
16  
17 modifies the sensitivity to these acoustic cues on both pitch and time dimensions.  
18  
19

20  
21 On a more linguistic side, Tyler and Cutler<sup>36</sup> compared English, French and Dutch speakers in  
22  
23 an artificial language learning paradigm and reported an interaction of participants' mother tongue and  
24  
25 the influence of vowel lengthening and pitch movement on speech segmentation performance.  
26  
27 Similarly, a speech stream respecting the native phonotactic rules results in better segmentation than a  
28  
29 "foreign" speech stream<sup>44</sup>. These manipulations of the stream could shed new light onto the extent to  
30  
31 which musical practice may affect language-universals and/or language-specific knowledge.  
32  
33

34  
35 Finally, a very promising direction will be to investigate learning processes per se, and to study  
36  
37 the familiarisation phase using electrophysiological measures (EEG/MEG). This approach has the  
38  
39 double advantage of an implicit measure of learning as well as of giving access to the learning curve  
40  
41 over time. Interestingly, some studies have reported different learning curves as a function of  
42  
43 behavioural performance, contrasting EEGs of high vs. low learners<sup>8,26,37,45</sup>. For the present aim,  
44  
45 advances in signal-processing methods, such as time-frequency analyses<sup>46</sup> and Frequency-tagging  
46  
47 analyses<sup>47</sup>, will allow for a better understanding of the cortical processes participating in speech  
48  
49 segmentation and will probably turn out to be highly informative when applied to investigate the bases  
50  
51 of the differences between musicians and nonmusicians in speech segmentation.  
52  
53  
54  
55  
56  
57  
58  
59  
60



## Acknowledgments

Preparation of this paper was supported by a grant from the French National Research Agency to D. Schön and B. Tillmann (ANR Blanc DMBB #NT09\_520631). C. François was PhD student supported by the ANR-Neuro (#024-01).

## References

1. Saffran, J. R., R. N. Aslin & E. L. Newport. 1996. Statistical learning by 8-month old infants. *Science*. 274 :1926-1928.
2. Kuhl, P. K. 2004. Early language acquisition: cracking the speech code. *Nature Review Neuroscience*. 5 :831-843.
3. Rodriguez-Fornells, A., T. Cunillera, A. Mestress-Misse & R. De Diego Balaguer. 2009. Neurophysiological mechanisms involved in language learning in adults. *Phil Trans of the Roy Soc B*. 364 :3711-3735.
4. Saffran, J. R., E. Johnson, R. Aslin & E. Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition*. 70(1) :27-52.
5. Aslin, R. N., J. R. Saffran & E. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*. 9(4) :321-324.
6. Gervain, J., F. Macagno, S. Cogoi, M. Peña & J. Mehler. 2008. The neonate brain detects speech structure. *Proc of Nat Acad of Sci*. 105 :14222-14227.
7. Teinonen, T., V. Fellman, R. Näätänen, P. Alku & Huotilainen, M. 2009. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*. 13 :10-21.
8. Abla, D., K. Katahira & Okanoya, K. 2008. On-line Assessment of Statistical Learning by Event related Potentials. *J of Cogn Neuro*. 20 (6) :952-964.
9. Kudo, N., Y. Nonada, N. Mizuno, K. Mizuno & K. Okanoya. 2011. On-line statistical segmentation of non-speech auditory stream in neonates as demonstrated by event-related brain



1  
2 potentials. *Developmental Science*. 14(5):1100-1106.

3  
4 10. Tillmann, B & S. McAdams. 2004. Implicit learning of musical timbre sequences: statistical  
5 regularities confronted with acoustical (dis)-similarities. *J Exp Psychol Learn Mem Cogn*. 30:1131-  
6  
7 1142.

8  
9  
10 11. Conway, M. C. & M. Christiansen. 2005. Modality-constrained statistical learning of tactile,  
11 visual, and auditory sequences. *J Exp Psychol Learn Mem Cogn*. 31(1):24-39.

12  
13 12. Hunt, R. H. & R. N. Aslin. 2001. Statistical learning in a Serial Reaction Time Task: Access  
14 to separable statistical cues by individual learners. *J Exp Psychol: General*. 130(4):658-680.

15  
16 13. Fiser, J. & R. N. Aslin. 2002. Statistical learning of higher-order temporal structure from  
17 visual shape sequences. *J Exp Psychol Learn Mem Cogn*. 28(3) :458-467.

18  
19 14. Fiser, J. & R. N. Aslin. 2005. Encoding Multielement Scenes: Statistical Learning of Visual  
20 Feature Hierarchies. *J Exp Psychol: General*. 134 (4):521-537.

21  
22 15. Schön, D., M. Boyer, S. Moreno, M. Besson, I. Peretz & R. Kolinsky. 2008. Song as an aid  
23 for language acquisition. *Cognition*. 106(2):975-983.

24  
25 16. François, C & D. Schön. 2010. Learning of musical and linguistic structures : comparing  
26 event-related potentials and behavior. *Neuroreport*. 21(14) :928-932.

27  
28 17. Sanders, L. D., E. L. Newport & H. J. Neville. 2002. Segmenting nonsense: an event-  
29 related-potential index of perceived onsets in continuous speech. *Nature Neuroscience*. 5(7) :700-703.

30  
31 18. François, C & D. Schön. 2011. Musical expertise boosts implicit learning of both musical  
32 and linguistic structures. *Cerebral Cortex*. 21(10) : 2357-2365.

33  
34 19. François, C. J. Chobert, M. Besson & D. Schön. Submitted. Music training for the  
35 development of language acquisition. *Cerebral Cortex*

36  
37 20. Hickok, G & D. Poeppel. 2007. The Cortical organization of speech processing. *Nature*  
38 *Reviews Neuroscience*. 8 :393-402.

39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
15

1  
2 21. Kraus, N. & B. Chandrasekaran. 2010. Music training for the development of auditory  
3 skills. *Nature Review Neuroscience*. 11(8):599-605.  
4

5  
6 22. Godey, B., D. Schwartz, J. B. De Graaf, P. Chauvel & C. Liégeois-Chauvel. 2001.  
7 Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a  
8 comparison of data in the same patients. *Clinical Neurophysiology*. 112 (10):1850-1859.  
9

10  
11 23. Molholm, S., S. Martinez, W. Ritter, D. C. Javitt & J. J. Foxe. 2005. The neural circuitry of  
12 pre-attentive auditory change detection an fMRI study of pitch and duration mismatch negativity  
13 generators. *Cerebral Cortex*. 15:545-551.  
14

15  
16 24. Saur, D., B. W. Kreher, S. Schnell, D. Kümmerer, P. Kellmeyer, M. S. Vry, R. Umarova, M.  
17 Musso, V. Glauche, S. Abel, W. Huber, M. Rijntjes, J. Hennig & C. Weiller. 2008. Ventral and dorsal  
18 pathways for language. *Proc of Nat Acad of Sci*. 105(46):18035-18040.  
19

20  
21 25. Flöel, A., M. H. De Vries, J. Scholz, C. Breitenstein & H. Johansen-Berg. 2009. White  
22 matter integrity in the vicinity of Broca's area predicts grammar learning success. *Neuroimage*.  
23 47(4):1974-1981.  
24

25  
26 26. Cunillera, T., E. Càmara, J. M. Toro, J. Marco-Pallares, N. Sebastián-Galles, H. Ortiz, J.  
27 Pujol & A. Rodríguez-Fornells. 2009. Time course and functional neuroanatomy of speech  
28 segmentation in adults. *Neuroimage*. 48 (3):541-553.  
29

30  
31 27. Sluming, V., T. Barrick, M. Howard, E. Cezayirli, A. Mayes & N. Roberts. 2002. Voxel-  
32 based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra  
33 musicians. *Neuroimage*. 17 (3):1613-1622.  
34

35  
36 28. Conway, C., D. Pisoni & W. Kronenberger. 2009. The Importance of Sound for Cognitive  
37 Sequencing Abilities: The Auditory Scaffolding Hypothesis. *Current Directions in Psychological*  
38 *Science*. 18(5):275-279.  
39

40  
41 29. Saffran, J.R., E. L. Newport & R. N. Aslin. 1996. Word segmentation: the role of  
42  
43  
44  
45  
46  
47  
48

1  
2 distributional cues. *Journal of Memory and Language*. 35:606-621.

3  
4 30. Perruchet, P. & S. Pacton. 2006. Implicit learning and statistical learning: one phenomenon,  
5  
6 two approaches. *Trends in Cognitive Science*. 10(5):233-238.

7  
8  
9 31. Perruchet, P. 2008. Implicit learning. In *Cognitive psychology of memory*. Vol.2 of *Learning*  
10  
11 and memory: A comprehensive reference (J. Byrne, Ed.) (pp. 597-621). Oxford: Elsevier.

12  
13  
14 32. Berry, D. C. & Z. Dienes. 1993. Implicit learning: Theoretical and empirical issues. *Essays*  
15  
16 in cognitive psychology. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

17  
18  
19 33. Tillmann, B. & B. Poulin-Charonnat. 2010. Auditory expectations for newly acquired  
20  
21 structures. *Quarterly Journal of Experimental Psychology*. 63:1646-1664.

22  
23  
24 34. Omigie, D. & L. Stewart. 2011. Preserved statistical learning of tonal and linguistic material  
25  
26 in congenital amusia. *Front. Psychology*. 2(109). doi: 10.3389/fpsyg.2011.00109.

27  
28  
29 35. Perruchet, P. & B. Tillmann. 2010. Exploiting multiple sources of information in learning an  
30  
31 artificial language: human data and modelling. *Cognitive Science*. 34:255-285.

32  
33  
34 36. Tyler, M. & A. Cutler. 2009. Cross-language differences in cue use for speech segmentation.  
35  
36 *Journal of the Acoustical Society of America*. 126:367-376.

37  
38  
39 37. De Diego Balaguer, R., J. M. Toro, A. Rodriguez-Fornells & A. C. Bachoud-Lévi. 2007.  
40  
41 Different Neurophysiological Mechanisms Underlying Word and Rule Extraction from Speech. *PLoS*  
42  
43 *ONE*. 2 (11):e1175.

44  
45  
46 38. Mirman, D., K. Graf Estes & J. S. Magnuson. 2010. Computational modeling of statistical  
47  
48 learning: Effects of transitional probability versus frequency and links to word learning. *Infancy*.  
49  
50 15:471-486.

51  
52  
53 39. Loui, P., D. L. Wessel and C. L. Hudson Kam. 2010. Humans rapidly learn grammatical  
54  
55 structure in a new musical scale. *Music perception*. 27(5):377-388.

56  
57 40. Reber, R & P. Perruchet. 2003. The use of control groups in artificial grammar learning.

1  
2 Quarterly Journal of Experimental Psychology. 56A(1):97-115.  
3

4 41. Perruchet, P., M. D. Tyler, N. Galland & R. Peereman. 2004. Learning nonadjacent  
5 dependencies: No need for algebraic-like computations. J Exp Psychol: General. 133(4):573-583.  
6

7  
8 42. Rohrmeier, M., P. Rebuschat & I. Cross. 2011. Incidental and online learning of melodic  
9 structure. Consciousness and Cognition. 20(2):214-222.  
10

11  
12 43. Cunillera, T., J. M. Toro, N. Sebastián-Gallés & A. Rodríguez-Fornells. 2006. The effects of  
13 stress and statistical cues on continuous speech segmentation: an event-related brain potential study.  
14 Brain Research. 1123 (1):168-178.  
15

16  
17 44. Cunillera, T., A. Gomila & A. Rodríguez-Fornells. 2008. Beneficial effects of word final  
18 stress in segmenting a new language: evidence from ERPs. BMC Neuroscience. 9:23.  
19

20  
21 45. Mersad. K & T. Nazzi. 2011. Transitional Probabilities and positional frequency  
22 phonotactics in a hierarchical model of speech segmentation. Memory and Cognition. 39(6):1085-1093.  
23

24  
25 46. De Diego Balaguer. R., L. Fuentemilla & A. Rodriguez-Fornells. 2011. Brain Dynamics  
26 sustaining rapid rule extraction from speech. Journal of Cognitive Neuroscience. 23(10):3105-3120.  
27

28  
29 47. Buiatti. M., M. Peña & G. Dehaene-Lambertz. 2009. Investigating the neural correlates of  
30 continuous speech computation with frequency-tagged neuroelectric responses. Neuroimage.  
31 44(2):509-519.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60