Human Beatboxing: A preliminary study on temporal reduction.

Alexis Dehais Underdown[1], Paul Vignes[1], Lise Crevier Buchman[1,2] & Didier Demolin[1]
[1] *Laboratoire de Phonétique et Phonologie (Sorbonne Nouvelle / CNRS),*
[2] *Hopital Foch, Univ. VSQ.*
alexis.dehais-underdown@sorbonne-nouvelle.fr

Abstract

This pilot study aims to understand the effect of speed rate on beatbox sounds. We carried out a temporal analysis on acoustic recordings of 11 Beatbox Patterns produced at 3 speeds: 90, 120 and 150 Beat Per Minute. Global results shows temporal reduction of patterns and sounds. Temporal reduction depends on the mode of articulation. However, for a same mode of articulation, sounds reduce in different ways. The position of sounds in a pattern is not correlated to temporal reduction. Although we noted temporal reduction, we only noted few omissions and beatboxing errors. Finally, we discuss the temporal organization of beatboxing gestures.

Keywords: Human Beatboxing, Temporal Reduction, Speed Rate.

1. Introduction

Speech rate is known to be a factor of reduction affecting supralaryngeal and laryngeal gestures. At fast rates, articulator velocity is increased and consequently gesture duration and amplitude are decreased (Lindblom 1963; Gay, et al.,1974; Ostry & Munhall, 1984; Munhall, & Löfqvist 1992; Byrd & Tan, 1996; van Son & Pol, 1999; Jannedy, Fuchs & Weirich, 2010). In addition, at fast rates, the interval of time between gestures is shortened and therefore the degree of overlap is increased (Munhall, & Löfqvist 1992; Byrd & Tan, 1996). Reduction is not linear, it depends on prosodic factors (Fougeron & Keating, 1997) and on the primary articulator (Gay, et al.,1974; Jannedy, Fuchs & Weirich, 2010). Little is known about acoustic reduction of consonants (van Son & Pol, 1999) and even less on the effect of beatboxing rate on production. Dehais Underdown et al. (2020) showed beatboxing rate affects the Center of Gravity (CoG) of hi-hats imitation but has little effects on the CoG of kick drums and snares. We do not know if rate has the same effects on speech tasks and beatboxing tasks because of the different nature of linguistic and musical systems. On one hand, speech must ensure efficient communication and depends on linguistic and communicative constraints. On the other hand, Human Beatboxing (HBB) is a form of artistic expression and relies on musical constraints. However, both speech and HBB are produced by means of vocal tract movements and thus are subject to physical constraints.

In the present study we are presenting an experiment based on a beatboxing speeding up task. We used a single metric and rhythmical pattern to create various beatbox patterns (BP). Since HBB is not linguistic, we wonder whether rate affects beatbox sound duration similar to speech. We expect that (1) the faster the production, the shorter sound duration (2) affricates, trills and fricatives will shorten more than stops (3) position in the pattern influences sound reduction.

2. Methods

A 32 y.o. professional beatboxer took part in the experiment. The recorded material was controlled in terms of sound pattern and rhythm. Indeed, to facilitate the comparison of beatbox recordings we used only one metric pattern that we transposed into 11 Beatbox Patterns (BP). Table 1 shows the 11 BPs created for the study. Concerning the rhythm, it was controlled using a vibrating metronome placed on his wrist. The participant was asked to repeat each BP 5 times in a row at 90 Beat per Minute (BPM), 120 BPM and 150 BPM (12 BP x 5 repetitions x 3 speeds = 180 BPs).

Recordings were manually segmented on Praat (Boersma & Weenink, 2006) in terms of acoustic phases (e.g. burst, friction, trilling), produced sounds and expected sounds (i.e. targets). To study temporal reduction, we extracted sound duration (ms),

Table 1 : *Metric, Rhythm and structure of Beatbox Patterns (BP). {ŋ [ ] ŋ} indicate sounds in these BP are velaric and the beatboxer can produce a melody while beatboxing.* indicates the pulsation.*

									
Positions	1*	2	3*	4	5	6*	7	8*	9
Structure of Beatbox Patterns of the corpus:									
1	[p’	t͡s’	↓k͡ʟ̥	t͡s’	p’	p’	t͡s’	↓k͡ʟ̥	t͡s’]
2	{ŋ [↑p	↑ts	k	↑ts	↑p	↑p	↑ts	k	↑ts] ŋ}
3	[t’	t͡s’	t͡ʃʀ̥’’	t͡s’	t’	t’	t͡s’	t͡ʃʀ̥’’	t͡s’]
4	[ɓ	t͡s’	↑ʔɛʃ	t͡s’	ɓ	ɓ	t͡s’	↑ʔɛʃ	t͡s’]
5	[↓p	t͡s’	p͡f’	t͡s’	↓p	↓p	t͡s’	p͡f’	t͡s’]
6	{ŋ [↑p	f	k	f	↑p	↑p	f	k	f] ŋ}
7	[p’	t͡s’	p͡f’	t͡s’	p’	p’	t͡s’	p͡f’	t͡s’]
8	{ŋ [↑p	↑ts	pf	↑ts	↑p	↑p	↑ts	pf	↑ts] ŋ}
9	[ɓ	t͡s’	↓iç	t͡s’	ɓ	ɓ	t͡s’	↓iç	t͡s’]
10	[↓ʙ̥ˡ	t͡s’	ʔəh	t͡s’	↓p	↓ʙ̥ˡ	t͡s’	ʔəh	t͡s’]
11	{ŋ [↓ʙ̥ˡ	↑ts	k	↑ts	↓ʙ̥ˡ	↓ʙ̥ˡ	↑ts	k	↑ts] ŋ}

acoustic phase duration (ms), silence between sounds (ms). To test the hypothesis of a possible relationship between the position and temporal reduction we plotted the duration of sounds and duration of BPs (i.e. speed rate index) and we indicated Pearson Coefficient for each position.

3. Results

3.1. Global effect of speed rate

Global results of speed rate shows duration of BPs is decreasing as the speed rate increases. We also observe that the faster the speed rate, the less standard deviation there is (Figure 1 error bars). Since BPs are composed of various type of sounds and silences between sounds, we need to have a closer look at how speed rate affect them.

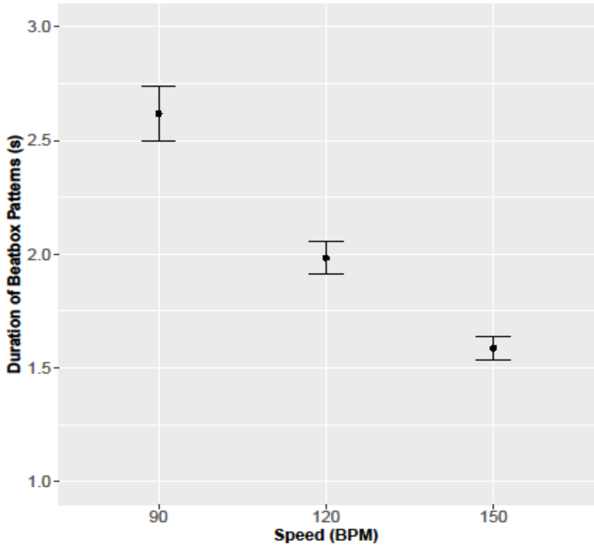

Figure 1: *Duration of Beatbox Patterns as a function of speed rate.*

Before turning to the results of sound reduction, it is important to mention that the beatboxer made few production errors. Indeed, when we compared the targeted sounds (i.e. the structure of BP to be repeated by the beatboxer) and the produced sounds (i.e. the structure of BP produced by the beatboxer) we found few omissions. On 1566 targets, only 22 (1.4%) were not produced at all; 18 of them were not produced at 150 BPM. We also noted 97 tokens (6%) that shows sound changes (e.g. [t͡s’] → [t͡ʃ’]; [↓p] → [↓ʙ̥ˡ]) but they will not be analyzed in this paper. In the following sections we removed all tokens that did not match the targeted sound. However we will discuss the interest of beatboxing errors in section 4.

3.2. Mode of articulation

We hypothesized that affricates, trills and fricatives will shorten more than stops. Figure 2 shows sound duration as a function of speed rate. Trills reduce most, then affricates and finally stops and fricatives were little reduced. It is important to note all sounds do not reduce the same way even if they share the same mode of articulation.

The voiceless ingressive pulmonic trill [↓ʙ̥ˡ] is longer than its velaric correlate {ŋ[↓ʙ̥ˡ]ŋ} (brackets indicate velaric initiation) and thus shortens more. Temporal differences between [↓ʙ̥ˡ] and {ŋ[↓ʙ̥ˡ]ŋ} are due to initiation mechanisms, the first being pulmonic ingressive and the latter being velaric ingressive. First, the velaric mechanism only modifies the pressure in the oral cavity unlike the pulmonic mechanism that modifies pressure in the entire vocal tract. Second, the velaric mechanism appears to be quicker to initiate the ingressive airflow required to create the differential of intraoral pressure (i.e. Intraoral pressure-Atmospheric pressure) and generate the trilling. Furthermore, we found two relevant parameters to describe trill reduction : the number of oscillations and the frequency of vibration. We noticed a reduction of the number of labial oscillation and an increase of the frequency of vibration at fast rate. In other word, lips are oscillating faster during a shorter interval of time.

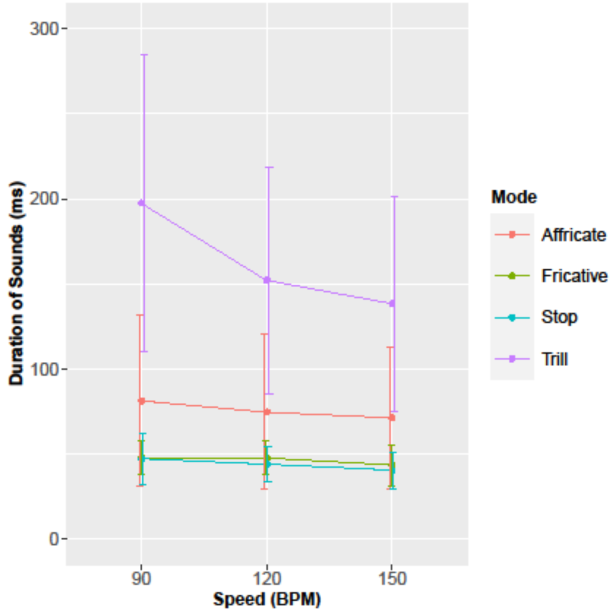

Figure 2: *Sound Duration (ms) as a function of speed rate. Each color indicates the mode of articulation.*

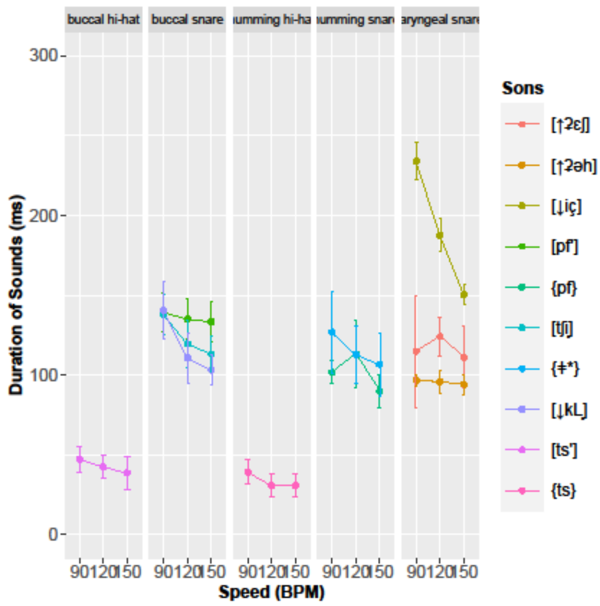

Figure 3: *Duration of affricates (ms) as a function of speed rate. Each color indicates the sound. Columns refer to the type of affricate (buccal, humming or laryngeal) and the type of instrument (hi-hat or snares)*
"Humming" is a beatboxing strategy relying on velaric initiation to allow simultaneous phonation.

Global analysis of duration for affricates revealed that they shorten in different ways. Detailed analysis of acoustic phases duration shows that affricates [↓k͡ʟ̥], [t͡ʃɪ̥’] as well as velaric snare {ŋ[k]ŋ} and hi-hats [ts’] and {ŋ[ts]ŋ} reduce the duration of the frication noise. However, the reduction of the labiodental affricates [pf’] (i.e. glottalic) and {ŋ[pf]ŋ} (i.e. velaric) seem to depend on the position. The laryngeal affricate are not affricates in a linguistic point of view. It refers to sounds composed of voiced onset, similar to a burst with vocalic resonances and a phase of frication or aspiration noise. [↓iç] highly reduces its frication noise whereas [↑ʔəh] and [↑ʔɛʃ] shorten only a bit. Ingressive are not common and difficult to study in speech (Eklund, 2008); but ingressiveness may explain why [↓iç] is highly reduced. We also found that the guttural snare [↑ʔəh] shortens in a particular way: its glottal onset (i.e. egressive voicing) lengthens but the aspiration noise shortens; though the overall duration shortens.
Finally, the velaric fricative {ŋ[f]ŋ} do not reduce as we thought, its reduction depends on the position. The analysis of the fricative is given in section 3.4.

3.3. Silences

BPs are characterized by an alternation of acoustic noise and silence (Figure 4). These silences are in fact intergestural intervals, that is, the time needed to move from one target to the next one. We annotated 2 different types of silence, (1) the silence between position 5&6 hereafter noted “-” (orange arrow) and (2) the rest noted “_” (blue arrows).

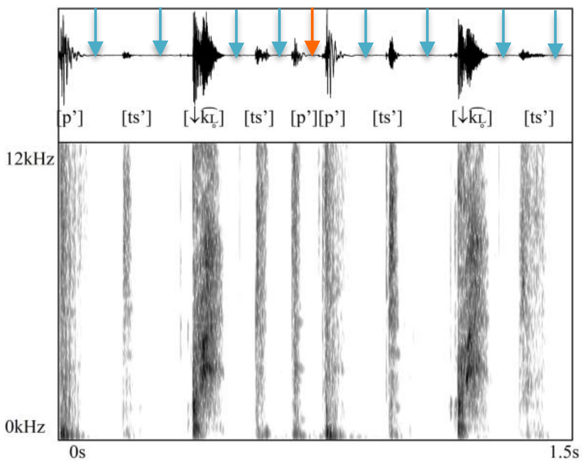

Figure 4: *Waveform and spectrogram of the BP [p’ ts’ ↓k͡ʟ̥ ts’ p’p’ ts’ ↓k͡ʟ̥ ts’] at 150 BPM.*

The first observation is that silences between position 5&6 (i.e. “-”) is shorter, because articulators hit the same articulatory target since sounds in position 5&6 share the same place of articulation. Figure 5 shows silence duration as a function of speed rate.
Silence reduction means gestures are getting closer on the time domain. It could lead to beatboxing errors such as assimilation, omission or transformation. Since the percentage of beatboxing omissions and transformations are low, we think the beatboxer is reducing both gestures’ duration but maintain the same relative overlap in order to avoid overlap.

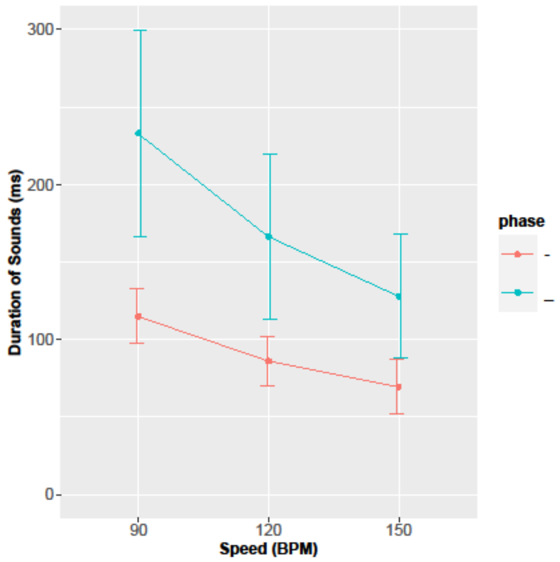

Figure 5: *Silence duration (ms) as a function of speed rate (BPM).*

3.4. Position

In section 3.2, we analyzed the effect of speech rate on sound duration regarding the mode of articulation. We will now have a look of the correlation between sound duration and speed rate for each position.

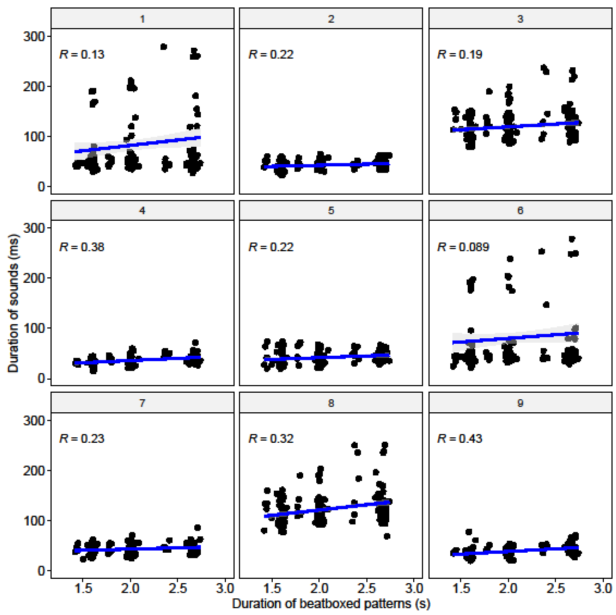

Figure 5 : *scatterplots of sound duration (ms) as a function of beatbox pattern duration (s) for each position*

Figure 5 shows the scatterplots of sound duration (ms) as a function of beatbox pattern duration (s) for each position. Since BPs decrease their duration at fast rate, we chose to plot their duration as an index of speed rate. Short durations correspond to fast rate (i.e. 150 BPM) and long durations correspond to low rate (i.e. 90 BPM). In order to evaluate the correlation between sound duration and BP duration we indicated the Pearson Coefficient. In position 1, 5 and 6, we kept stops (i.e. kick drums) and trills (i.e. lips roll) together even if trills are longer.

Removing trills from scatterplots barely changed *R* values from 0.13, 0.22, 0.089 to 0,17, 0.19, 0.23 respectively in position 1, 5 and 6. Even though the coefficient in position 6 is higher, the correlation is still weak.

No correlations were found in the current data. Three positions show higher *R* coefficient. Position 4 ($R = 0.38$) corresponds to the end of the first half of the pattern; it is also the position where most omissions and transformation occurred. Final positions also show higher *R* coefficients : position 8 $R = 0.32$ and position 9 $R = 0.43$. Higher *R* coefficients means more reduction in these specific positions. It is interesting to note reduction is more important in final position. However, the correlation is weak and our hypothesis cannot be confirmed.

We found two interesting cases where sounds lengthen as speed increase. The voiced kick [ɓ] lengthens in position 5 whereas it shortens in position 6. Our corpus is built so the same sound appears in position 5&6. In absence of physiological data we cannot evaluate possible constraints or explanation of lengthening. The velaric hi-hat {ŋ[f]ŋ} lengthens in position 2 and 7 but shortens in position 4 and 9. Velaric fricatives are not attested in phonological systems and we do not know yet how it is produced. Velaric mechanism cannot explain the lengthening on its own, otherwise we would have noticed similar patterns of reduction/lengthening in other velaric sounds. However, the combination of a velaric mechanism and a fricative gesture could be one part of the explanation. The other part would be the effect of segmental context. Position 2 and 7 have the same segmental context: {ŋ[↑p **f** k]ŋ}; whereas position 4 and 9 have a different context: {ŋ[k **f** ↑p]ŋ} or {ŋ[k **f** #]ŋ}. A labial constraint between {ŋ [↑p] ŋ} and {ŋ[f] ŋ} might provoke the lengthening.

4. Discussion

The data indicates that the beatboxer adapted to the speeding up task with precision: there is evidence of temporal reduction but few beatboxing errors (i.e. 1.4% of omissions and 6% of transformations). We validated the first two hypothesis: (1) the faster the production, the shorter sound duration (2) affricates, trills and fricatives will shorten more than stops. Speed rate has a global effect on the beatboxer's production, that is a reduction of the duration of Beatbox Pattern. A more detailed analysis on sound duration revealed reduction depends on the mode of articulation. For a same mode of articulation, sounds do not always reduce the same way. Concerning the second hypothesis we found evidence that trills reduce most and then affricates > fricative-stop. Aerodynamic data is needed to understand the interplay between initiation mechanisms, airflow, intraoral pressure and reduction. It could help to find, or not, a clearer pattern of reduction. The last hypothesis is not confirmed; the position of sounds in the pattern is not correlated to temporal reduction for the BPs analyzed in the present study. Yet, the absence of correlation do not imply rhythmical structure has no effect on reduction. Indeed, the absence of correlation may be an artefact of our corpus. A temporal analysis of freestyle recordings could provide evidence of a relationship between rhythm and reduction.

Beside the previous findings, two interesting findings drew our attention: gestures are getting closer at fast rate and there are only 6% of transformations. These transformations seem to be cases of overlap. We found cues of labialization of [ts'] → [tsf'] when it preceded a labial or was between two labial sounds in two specific BPs (1) [p' ts' ↓k͡ʟ̥ ts' p'p' ts' ↓k͡ʟ̥ **ts'**] and (2) [p' ts' pf' **ts'** p'p' ts' pf' ts']. In the first BP there were only two tokens of [tsf'] found in position 9 at 150BPM. Since there are 5 repetitions of BPs, labialization must be related to the first kick of the following pattern. For the second BP, where [ts'] was between two labial sounds, we found 8 tokens of labialization at all rates but only in position 4. This suggests evidence of anticipatory movement of the jaw. Beatboxing errors need to be included in further analysis.

Beatboxing errors, however, are somehow scarce, even if gestures are closer to one another. The temporal organization of beatboxing gestures appears to be different than speech gestures. It is possible that beatboxing gestures are not planned with overlap contrary to speech where syllable structure favors overlap. In HBB there are no "vowels" (Figure 4), at least not in our data, and thus there is no CV structure. In this sense, HBB is quite different from speech. The BPs to be repeated have a simple structure, more complex structures could mean a different way of organizing gestures and overlap.

5. Conclusion

This study proposes the first temporal analysis of the effect of speed rate on Human Beatboxing. Speed rate has a different effect on production in HBB. It is a rich paradigm to study articulators motion and coordination in a non-linguistic task. Playing the game "spot the differences" between speech and HBB could provide useful comprehension of production mechanisms.

6. Acknowledgements

We thank the beatboxer for his time and precious collaboration and to have made the study possible.

7. References

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer. Version 6.0.21, recupéré le 25 Septembre 2016 sur http://www.praat.org/.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24(2)*, 263-282.

Dehais Underdown, A., Vignes, P., Buchman, L. C., & Demolin, D. (2020). Rythme et contrôle articulatoire: étude préliminaire du Human Beatbox. In *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 1: Journées d'Études sur la Parole* (pp. 136-144). ATALA.

Eklund, R. (2008). Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association*, 235-324.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The journal of the acoustical society of America, 101(6)*, 3728-3740.

Gay, T., Ushijima, T., Hiroset, H., & Cooper, F. S. (1974). Effect of speaking rate on labial consonant-vowel articulation. Journal of *Phonetics, 2(1)*, 47-63.

Jannedy, S., Fuchs, S., & Weirich, M. (2010). Articulation beyond the usual: Evaluating the fastest German speaker under laboratory conditions. *Between the regular and the particular in speech and language*, 205-234.

Lindblom, B. (1963). Spectroraphic Study of Vowel Reduction. *Journal of the acoustical society of America*, 1773-1781.

Munhall, K., & Löfqvist, A. (1992). Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics, 20(1)*, 111-126.

Ostry, D. J., & Munhall, K. G. (1985). Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America, 77(2)*, 640-648.

Van Son, R. J. J. H., & Pols, L. C. (1999). An acoustic description of consonant reduction. Speech communication, 28(2), 125-140.