

Cantar junto em dois tipos de ritmo: Um estudo preliminar

CÁSSIO AUGUSTO SANTOS* E BEATRIZ RAPOSO DE MEDEIROS**

Resumo

Este é um estudo experimental acerca da fala cantada no português brasileiro, ou simplesmente do canto tal qual o conhecemos na canção popular brasileira. Tendo como objetivo observar o processo de sincronização entre duplas de cantores, elaboramos um experimento com dois fatores: Ritmo da Canção e Ponto de Observação de Sincronização. A partir do fator Ritmo da Canção buscamos observar se o processo de sincronização na fala cantada é influenciado pelo ritmo musical. A partir do fator Ponto de Observação de Sincronização buscamos observar se os sujeitos têm a mesma capacidade de sincronização em inícios e em finais de frases. Para quantificar a sincronização do canto entre sujeitos cantores de uma mesma dupla, realizamos medidas das *lags*, ou seja, a diferença temporal entre dois pontos do sinal sonoro, que chamamos de Pontos de Observação de Sincronização. A sincronização mostrou-se uma tarefa mais fácil em ritmo simples (não sincopado), que foi o caso das canções de *rock* e apresentou *lags* mais longas, tanto no início como no final da frase (pontos diferentes observados) no ritmo complexo das canções da bossa nova. Este resultado indica uma interação das capacidades rítmicas da fala e da música cuja discussão pode servir de base para demais estudos sobre a relação canto e fala.

Palavras-chave: ritmo, canto, fala, sincronização

Singing together in two types of rhythm: A preliminary study

Abstract

This paper presents a preliminary experimental study about synchronized singing in Brazilian Portuguese as it is commonly known in the Brazilian popular song. Aiming to observe a synchronization process between a dyad of singers, this study presents experiment factorial design with two factors, called Rhythm of Song and Points of Note Synchronization. With the Rhythm of Song factor, we sought to observe the synchronization process in singing within the rhythmic-musical context. With factor Points of Note Synchronization we sought to determine if the subjects have the same ability to synchronize at the beginning and end of sentences. In order to quantify synchronization between two singers in a dyad, we measured the lag duration, it is to say, we measured the time difference between two points in the sound wave, i.e., the Points of Note Synchronization. Synchronization showed to be an easy task in a simple rhythm (non-syncopation), that was the rock songs case, and showed longer lags both in the initial and final sentence point in the complex rhythm. This result points to an interplay between rhythmic abilities both in speech and music which discussion can ground further studies on singing and speech relation.

Keywords: rhythm, song, speech, synchronization

* Universidade de São Paulo - USP
E-mail: cassioandradesantos@hotmail.com

** Universidade de São Paulo - USP
E-mail: beatrizrap@gmail.com

Recebido em 3 de outubro de 2015; aceito em 7 de dezembro de 2015.

1. Introdução

Apresentamos um estudo experimental de caráter preliminar que buscou observar como se dá a sincronização entre cantores na realização de dois ritmos cancionais diferentes, e bem familiares na cultura musical brasileira: a bossa nova e o *rock*. O canto com texto, que aqui chamamos de fala cantada, é, pois, nosso objeto de estudo. O termo “fala cantada” foi empregado para o estudo comparativo sobre diferenças e semelhanças de aspectos fonéticos entre canto e fala (Raposo de Medeiros, 2002). Uma vez que todos os aspectos acústicos da fala eram vistos tanto no canto como na fala foi necessário diferenciar os dois fenômenos em comparação, levando-se em conta o que tinham em comum: “fala falada” (fala) da “fala cantada” (canto). No caso do presente estudo, a fala cantada observada é realizada em duas condições rítmicas diferentes, que chamamos, até o presente, de sincopada (bossa nova) e não sincopada (*rock*).

A ideia de desenvolver um estudo sobre sincronização da fala cantada surgiu de questionamentos mais gerais sobre a influência da estrutura temporal musical sobre a fala e o que essa influência poderia revelar sobre o aspecto rítmico dessa última. Um exemplo corriqueiro disto é o deslocamento de acento em uma palavra, de uma sílaba para outra, a fim de atender aos acentos musicais. Por si, esse exemplo mostra como a fala pode ser elástica e aceitar alguns deslocamentos sem comprometer o texto da canção—ou seja, sua parte linguística. À luz da ideia de elasticidade da fala e inspirados pelos primeiros trabalhos de Fred Cummins (2002, 2003, 2009) sobre a sincronização da fala e sua ideia mais geral sobre ritmo, elaboramos um experimento sobre a fala cantada.

Cabe ressaltar, então, que o achado de Cummins (2003), ou seja, de que dois falantes sincronizam suas falas sem maiores problemas e sem treino, tem importância crucial para nosso entendimento de ritmo na fala. Trata-se de entender o fenômeno da sincronização como resultante da capacidade de ajustar movimentos, que, no caso da fala, são os movimentos de seus articuladores. Transfere-se, pois, a preocupação excessiva em se explicar os aspectos rítmicos da fala a partir de pistas acústicas e da ideia de isocronia para a ideia de ajuste dos movimentos. Este ajuste ou integração de movimentos é chamado de *entrainment*, processo sobre o qual falaremos com detalhes mais adiante.

Em um primeiro momento, nossa proposta de que diferentes ritmos podem se sincronizar de diferentes maneiras (ver seção 4) pode parecer não alinhada à ideia de que a fala se sincroniza facilmente, independente de não oferecer pistas, como batidas isócronas—que

seriam uma espécie de grade métrica. No entanto, nossa proposta é de caráter preliminar e diferencia-se dos estudos de Cummins pelo fato de, na fala cantada, a estrutura temporal musical abrigar a estrutura temporal da fala, sem que esta última perca suas características. Assim, a intenção é observar a sincronização da fala cantada em condições rítmicas da música que, pressupomos, tenham influências diferentes na integração de dois domínios diferentes, o da fala e o da música, que, juntos, formam o domínio da canção.

2. Conceitos tradicionais de ritmo da fala

Ao longo de várias décadas, estudos voltados para a compreensão do ritmo da fala (Abercrombie, 1967; Lloyd-James, 1940¹, como citado em Barbosa, 2006; Pike, 1945) buscaram relacionar diferentes aspectos desta, como acento e duração silábica, a duas classificações rítmicas: o ritmo acentual e o ritmo silábico. Embora estas classificações já tenham sofrido críticas, por se basearem na noção de isocronia entre sílabas ou acentos (Dauer, 1983; Cummins, 2002), são recorrentes em análises e tipificação rítmica das línguas naturais. A literatura a respeito busca classificar as línguas da mesma maneira como se classificam os ritmos musicais² (samba ou *rock*, por exemplo), separando-as, neste caso, categoricamente. No entanto, autores já citados das críticas aos rótulos rítmicos silábica e/ou acentual, defendem que tal distinção categórica não corresponde à realidade do que é produzido ritmicamente na produção da fala.

Na produção da fala cantada, a articulação do trato vocal sofre algumas modificações justamente devido a outra estrutura que também sustenta a canção que não a linguística: a estrutura musical (Raposo de Medeiros, 2002). Os mesmos fones usados na fala falada são transportados ao universo da música, universo dominado por frequências pré-determinadas (as notas musicais) e por relações de tempo (duração e ritmo) que acabam por caracterizar a fala cantada como tal, exigindo manobras articulatórias diferentes daquelas exigidas pela fala falada

Inserir a fala no contexto musical é transportá-la para uma estrutura temporal bem mais rígida que a da fala falada, pois o tempo na música é o que move³ uma peça e o que, justamente, ajuda músicos de uma mesma orquestra a tocarem ou a cantarem “a mesma coisa ao mesmo tempo”.

¹ Lloyd-James, A. *Speech signals in telephony*. Londres: Sir Isaac Pitman and Sons, 1940.

² Acreditamos que classificar as línguas do mundo nesta dicotomia seria como tentar classificar todas as canções do mundo na dicotomia samba *versus* rock, por exemplo, ou seja, ignorar outros ritmos e suas características. Tal dicotomia, também, leva a entender que os ritmos não conseguem dialogar entre si, sendo assim tipos com formas rígidas e sem interação com outros.

³ A palavra ritmo e outras palavras a ela relacionadas, como troqueu, marcha e *rock and roll*, está etimologicamente relacionada à ideia de movimento.

3. Novas propostas para se entender o ritmo linguístico

Partindo do ponto de vista de que falar é movimento, Cummins (2009b) apresenta uma nova proposta de ritmo aplicada às línguas, e mais do que isso, aplicada não só às línguas, mas também à música, à dança e a qualquer fenômeno ou processo nos quais movimentos, de um modo geral, são observados a partir de um ponto de vista dinâmico, ou seja, um ponto de vista em que as mudanças sofridas por um sistema ao longo do tempo são importantes para explicar este mesmo movimento. Para esse autor, ritmo é viabilizador do carreamento do movimento (*an affordance for the entrainment of movement*), uma ideia que necessita de alguns esclarecimentos conceituais para ser entendida.

O conceito de *affordance* (Gibson, 1979; Chemero, 2003; como citados em Cummins, 2009b⁴) é oriundo da psicologia ecológica e define “*relations between the abilities of organisms and features of the environment*” (p. 5). Como exemplo, o autor traz a relação de “agarrabilidade” entre um copo e uma pessoa que busca agarrá-lo: essa relação só existe se o copo for “agarrável” (se tiver uma alça ou se não estiver muito quente, por exemplo) e se a pessoa for capaz de agarrá-lo (uma pessoa sem mãos não conseguiria realizar tal ação ou uma pessoa sem luvas não poderia agarrar um copo a determinada temperatura). Ora, aplicado ao estudo do ritmo das línguas, *affordance* define a relação entre as habilidades de movimento de fala de um sujeito e a estrutura temporal de sua língua (Cummins, 2009b).

Entrainment, que aqui podemos traduzir por carreamento⁵ é um conceito trazido da física e das ciências biológicas que lidam com o tempo. É definido como o processo no qual dois osciladores interagem entre si de modo a se ajustarem e, eventualmente, entrarem em fase ou período⁶ (Clayton, Sager, & Will, 2004). A sincronização aqui observada advém do *entrainment* entre dois cantores ou cantoras cantando simultaneamente a mesma canção. A *sincronia* perfeita (com fase igual a zero, chamada de *zero phase*) seria uma das consequências do *entrainment* e assim distinguimos sincronização como um processo; processo esse que pode, como consequência, implicar em uma sincronia perfeita ou não entre os osciladores envolvidos.

⁴ Gibson, J. J. (1979). *The ecological approach to visual perception*. (Houghton Mifflin, Boston, 1979). Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, 15(2), 181–195.

⁵ O termo “carreamento” foi inicialmente usado pela linguista Eleonora Albano e recupera, do inglês, o sentido de engatar os vagões de trem entre si, sentido que está presente na palavra em inglês *entrainment*.

⁶ [in.train.m'nt] se difere de [en.treyn.ment], ambos grafados da mesma forma e derivados da palavra *train* (trem). O primeiro se refere ao processo acima referido. O segundo é um conceito usado na físico-química e na meteorologia e se refere ao arrastamento de certos gases por uma nuvem de vapor ou pelo arrastamento de um fluido por outro.

Port, Tajima e Cummins (1996) relatam os resultados de algumas experiências que testam a relação entre inícios de frases repetidas e a posição das sílabas tônicas destas frases em diferentes andamentos. Eles relatam que seres humanos realizam maior ou menor sincronização na fala, dependendo da condição da tarefa e sugerem ainda que estudos sobre autossincronização (sincronização de gestos articulatórios da fala de um mesmo indivíduo, ou sincronização com outros gestos que não apenas os gestos da fala) podem ser profundamente reveladores sobre a maneira em que o tempo é tratado no sistema nervoso dos animais, indo muito além da coordenação dos próprios membros, como braços e pernas ao andar. Mayberry, Jaques e Dede (1998⁷), por exemplo, observaram que pessoas que sofrem de gagueira tendem a parar o movimento das mãos ao gaguejar, o que exemplifica o processo de autossincronização entre a fala e um outro sistema de movimento, neste caso, as mãos. Na realização da fala, ainda, nota-se o processo de autossincronização quando o falante coordena os movimentos dos diferentes órgãos articulatórios, como o arredondamento dos lábios que se coordena ao movimento da língua na realização da sílaba /du/, por exemplo. Ao produzirmos esta sílaba, percebemos que os lábios já se arredondam quando articulamos o início do [d], ao mesmo tempo em que a ponta da língua se posiciona para a articulação característica deste fone, que é a do seu contato com os dentes superiores.

Cummins (2002, 2003, 2006, 2009a, 2009b) realizou diversos trabalhos utilizando o processo de sincronização na fala como ferramenta para estudar o ritmo da fala como fenômeno não isócrono, mas que permite a sincronização. Em um de seus experimentos, Cummins (2002), mediu a sincronização entre dois falantes instruídos a ler um texto juntos e comparou aos resultados obtidos quando um desses falantes lia o mesmo texto acompanhando uma gravação realizada pelo parceiro da primeira tarefa. O autor concluiu que a segunda tarefa é mais difícil de ser realizada (medianas de 30 milissegundos na primeira tarefa contra 56 milissegundos na segunda). Cummins (2003) observou também a sincronização em três diferentes pontos das sentenças de um texto— inicial, medial e final—e notou menor sincronia no ponto inicial, não observando diferenças entre a sincronia nos pontos finais comparados aos mediais. Observou-se também que a prática não tem grande influência na melhora da realização da tarefa, ou seja, que os falantes têm uma grande capacidade de sincronização, mesmo

⁷ Mayberry, R. I., Jaques, J., & Dede, G. (1998). What stuttering reveals about the development of the gesture-speech relationship. In J. M. Iverson; Goldin-Meadow, S. (Eds.), *The Nature and Function of Gestures in Children's Communication. New Directions for Child Development*, no.79. San Francisco: Jossey-Bass.

sem prática. Nesse trabalho, o autor notou, ainda, a importância do contato visual entre os falantes, afirmando que a tarefa é mais fácil de ser realizada com a manutenção deste tipo de contato e que os sujeitos fazem, quase que exclusivamente, uso da visão periférica ao invés da visão foveal (a visão que projeta a imagem do objeto focalizado). Finalmente, ainda concluindo o observado no estudo de 2002, Cummins (2003) notou que a boa realização da tarefa de sincronização depende da colaboração mútua entre os falantes, ou seja, na fala sincronizada não há um líder, os dois falantes interagem de forma a colaborarem mutuamente.

4. Hipóteses

Como já dito anteriormente, este trabalho apresenta um estudo experimental do ritmo da fala cantada à luz do conceito de *entrainment* (Cummins 2009; Clayton, Sager, & Will, 2004) e com base nos trabalhos experimentais de Cummins (2002, 2003, 2006, 2009).

Inicialmente colocamos duas hipóteses principais. A primeira defende a ideia de que seja mais fácil para os sujeitos sincronizarem suas falas em canções de ritmo não sincopado, como é o caso do *rock*. Mesmo que a linha melódica do *rock* possa ter figuras rítmicas sincopadas, seu acompanhamento é sempre um ritmo mais simples, diferente do samba, que oferece sincopação tanto na melodia, quanto nos vários tipos de acompanhamento. Acreditamos que a síncope, conflito métrico-rítmico, tenha influência na realização da tarefa da sincronização, podendo dificultá-la. A segunda hipótese, baseada em estudos de Cummins (2002, 2003, 2006), defende a ideia de que, independentemente do ritmo, as regiões finais das sentenças seja mais fácil de sincronizar quando comparadas às regiões iniciais das mesmas. Estas hipóteses germinaram em searas diferentes: os fatores do experimento realizado.

A partir da combinação entre os fatores pudemos fazer emergir uma terceira hipótese: que seja mais fácil sincronizar a fala em pontos finais de sentenças não sincopadas.

5. Metodologia

5.1 Delineamento experimental

Desenvolvemos um experimento no qual pudéssemos observar o processo de sincronização na fala cantada em dois diferentes contextos rítmico-musicais: o *rock* e a bossa-nova; e em dois pontos diferentes das sentenças do *corpus*: ponto inicial e ponto final. A partir da visua-

lização da onda sonora por meio de um *software* de análise acústica⁸, medimos pontos de sincronização entre os cantores de uma mesma dupla, que cantaram as canções em uníssono.

O ritmo das canções e os pontos de observação da sincronização formam, cada um, um “fator experimental” de nosso experimento, definido, portanto, como um “experimento fatorial com dois fatores” (Calegare, 2001). O fator A foi denominado “Ritmo da Canção” e o fator B, “Ponto de Observação de Sincronização”. Por sua vez, cada um desses fatores foi dividido em dois níveis: o fator A apresenta os níveis Ritmo Sincopado (RS) e Ritmo Não-Sincopado (RNS); enquanto o fator B apresenta os níveis Ponto Inicial (PI) e Ponto Final (PF), como ilustrado na *tabela 1*:

Tabela 1: Delineamento do experimento. Os fatores e seus respectivos níveis.

Fatores	Ritmo da Canção		Ponto de Sincronia	
Níveis	RS	RNS	PI	PF

As quatro canções selecionadas, compostas por cancionistas brasileiros, são: *Águas de Março* e *Corcovado*, ambas de Antônio Carlos Jobim; *Aluga-se*, de Raul Seixas e Cláudio Roberto e *Gita*, de Raul Seixas e Paulo Coelho. As duas primeiras foram agrupadas no nível Ritmo Sincopado, enquanto as duas últimas no nível Ritmo Não-Sincopado, já que, por definição, a bossa-nova possui o rítmico de acompanhamento/grade métrica sincopado, o que não é o caso do *rock*.

Como apresentaremos adiante mais detalhadamente, no fator “Ponto de Observação de Sincronização” colocamos as quatro canções, porém divididas em sentenças, nas quais colocamos duas marcações: um ponto inicial e um ponto final.

5.2 Coleta dos dados

Para a coleta dos dados, contamos com quatro diferentes duplas de sujeitos, todas formadas por cantores e cantoras profissionais com idades entre 22 e 32 anos e com mais de três anos de experiência com “canto popular”. O *set* de gravação foi disposto de forma que os cantores ficassem a cerca de três metros um do outro e no mesmo ambiente de gravação, ou seja, com contato visual e sonoro direto. Utilizou-se o *software Sony Soundforge* para coleta/digitalização do áudio, salvos cada um em um canal e em formato WAV.

Como tarefa, pedimos às duplas que cantassem as canções selecionadas em uníssono, de forma mais sincronizada possível, *a capella*

⁸ O programa utilizado foi o *Praat*, programa gratuito desenvolvido por Paul Boersma e David Weenink: <http://www.fon.hum.uva.nl/praat/>

e na tonalidade e ordem que preferissem. Cada canção foi cantada três vezes e a essas vezes precederam algumas, a fim de adaptação ao ambiente do estúdio, alongamento e aquecimento vocais, sendo que as três repetições de cada canção foram coletadas em sequência. Passamos a gravar quando as duplas se declararam prontas para tal. De acordo com Cummins (2003), a partir da terceira repetição não há diferenças significativas na tarefa de sincronização e por isso pedimos aos sujeitos que cantassem as canções algumas vezes antes da gravação e passamos a gravar, pelo menos, a partir da terceira.

Do que foi exposto até aqui, desde a descrição dos sujeitos, até a da coleta, reiteramos que buscamos montar duplas homogêneas, com sujeitos que realizassem a tarefa com igual nível de exigência e de desenvoltura. Os sujeitos puderam cantar as canções na tonalidade de sua tessitura, todos já conheciam as canções e todos eles foram instruídos da mesma forma. Todos eram cantores profissionais, tendo passado por estudo formal do canto popular e nenhum deles era intérprete só de *rock* ou só de *bossa nova*.

Em nosso desenho experimental, a ideia de uma dupla de cantores é um sistema de dois osciladores acoplados, no qual cada um dos sujeitos é um oscilador. Eles realizam movimentos, os movimentos próprios da fala cantada, a partir de (i) fontes de energia interna próprias (o sistema respiratório) e (ii) tais movimentos são ao mesmo tempo influenciadores e influenciados pelos movimentos do outro oscilador do sistema: o outro cantor, corroborando, assim, com os conceitos de oscilador e de osciladores acoplados⁹ (Clayton, Sager, & Will, 2004).

O ritmo que observamos aqui é o elo que permite a comunicação motora e articulatória entre os dois sujeitos, permitindo-lhes que se movimentem de forma sincrônica, ou a mais sincrônica possível. Um ritmo não-sincopado, portanto aquele que não entra em conflito com a estrutura métrica básica, é um “elo” mais simples, que facilita a coordenação dos movimentos justamente pela ausência da síncopa, do conflito. Em comparação a este, o ritmo sincopado é um elo de maior complexidade, justamente devido à presença da síncopa, que acreditamos tornar tal elo mais frágil, mais fácil de ser rompido.

5.3 Segmentação e definição dos pontos para as medidas

Segmentamos as canções em sentenças, definidas a partir do conceito de frase musical (Schönberg, 1996). Em seguida, rotulamos essas sentenças em duas regiões diferentes — inicial e final — nas quais obser-

⁹ A inserção de um metrônomo, por exemplo, que guiasse os cantores na tarefa, iria contra o conceito de oscilador, já que o metrônomo não sofre influência dos cantores.

vamos a sincronização entre os sujeitos de cada dupla. A partir desta rotulagem, conseguimos medir a diferença da sincronização entre os dois cantores, diferença essa que chamamos de *lag* (para ilustração do *lag*, ver *Figura 1*). A *lag* consiste na diferença de início (*onset* ou ataque) de cada ponto pré-estabelecido do sinal sonoro observado, ou seja, as frases cantadas.

A rotulagem dos pontos foi feita com a colocação de uma marcação em algum dos primeiros fones de cada sentença, ao qual chamamos de “Ponto Inicial”; e outra marcação colocada em algum dos últimos fones de cada sentença, a qual chamamos de “Ponto Final”.

Obtivemos 44 sentenças da canção *Gita* e 20 da canção *Aluga-se*, somando 64 na condição RNS. Da canção *Corcovado* obtivemos 12 sentenças e de *Águas de Março*, 51, somando 63 na condição RS.

No entanto, algumas destas sentenças não se encaixavam em nosso delineamento, ora por serem muito curtas, ora por não apresentarem fones oclusivos ou fricativos. Essas frases foram excluídas das análises. Outro motivo que nos levou a retirar algumas outras sentenças foi o fato de querermos a mesma quantidade de sentenças nas duas condições, a fim de simplificar as análises estatísticas. Assim, conseguimos obter 55 sentenças em cada condição.

Multiplicando-se esse valor, 55, pelo número de repetições, três, pela quantidade de duplas, quatro, e pela quantidade de pontos de observação de sincronização em cada sentença, dois, obtêm-se 1320 *lags* para cada nível do fator Ritmo da Canção. Multiplicando-se esse resultado pela quantidade de níveis do fator, dois, obtêm-se 2640 *lags* no total. O mesmo cálculo pode ser feito com o fator Ponto de Observação de Sincronização¹⁰.

Em cada uma das sentenças obtidas, colocamos dois pontos pelos quais observamos a sincronização entre os cantores. Esses pontos, um no início (Ponto Inicial) e outro no final (Ponto Final), foram colocados principalmente a partir de critérios fonéticos, mas também a partir de critérios musicais¹¹. Priorizamos a colocação desses pontos no ataque de determinados fones fricativos e oclusivos por terem a borda esquerda mais facilmente distinguível e cujo *onset* coincidia com o ataque de notas musicais. Não conseguimos fazer assim com todos os pontos, devido à natureza das canções, fato que nos levou a colocar alguns deles

¹⁰ Pede-se atenção para o fato de o fator Ponto de Observação de Sincronização não ser composto de outras 2640 *lags*, mas de uma diferente distribuição das mesmas, ou seja, ora as *lags* foram separadas por ritmo, pela classificação “sincopado” e “não sincopado”, ora foram separadas por ponto de observação de sincronização, pela classificação “inicial” e “final”.

¹¹ Utilizamos este método por ser o mais viável no momento deste estudo. Em estudos mais recentes o autor passou a utilizar outra metodologia, na qual é realizada a medição de uma quantidade maior de pontos por sentença, pontos estes determinados por programas de computador, a partir da identificação de fonemas vozeados.

no *onset* de outros fonemas que não oclusivos ou fricativos (mas que mesmo assim tinham o início bem definidos—como vogais em início) e cujos inícios coincidiam ainda com o ataque de notas musicais. Foram poucos esses casos.

A localização dos pontos em fones assim caracterizados foi importante para a precisão da ordem de milissegundos ao determinar as *lags*, e conseqüentemente, a sincronização entre os pontos. Colocar um ponto de observação de sincronização, por exemplo, entre duas vogais, seria tarefa difícil, já que esses gestos vocálicos podem se sobrepor de tal forma que seria difícil identificar o *onset* do segundo fone, onde seria colocado o ponto.

Colocados os pontos no sinal acústico dos dois cantores, sobrepostos os dois rótulos, podendo calcular a diferença entre os pontos de observação de sincronização, pois o *Praat* nos fornece a posição numérica dos pontos no eixo do tempo:



Figura 1: Pontos Inicial e Final da cantora LS comparados aos pontos Inicial e Final de HC.¹²

A *figura 1* mostra, portanto, quatro pontos: dois iniciais e dois finais, da primeira sentença da canção *Aluga-se*, cantada pela dupla LS e HC. A linha 1 mostra os pontos colocados no sinal acústico da cantora LS e a linha 2, os pontos colocados no sinal acústico do cantor HC. A sincronização entre os cantores pode ser vista, portanto, nas *lags*: a diferença temporal (representada na figura através da diferença espacial) entre os pontos respectivos entre si, como nos exemplos circulares em amarelo (*Pontos Iniciais*) e em verde (*Pontos Finais*). Essas *lags* são, de fato, aquilo que tomamos nesse experimento como dado, pois a partir delas, quantificamos a sincronização.

¹² Sobrepostos os rótulos dos dois cantores, o *Praat* só apresenta o espectrograma de um deles. Na *figura 1*, vê-se o do cantor HC.

6. Resultados

Inicialmente, apresentamos a estatística descritiva geral dos dados, o que servirá de base para a comparação entre os fatores e seus níveis. A tabela a seguir traz, portanto, a média e o desvio-padrão de todos os dados, além da mediana e dos valores máximo e mínimo¹³ de *lags* encontrada, todos em milissegundos (doravante ms):

Tabela 2: Estatística dos dados gerais. Valores das *lags* em ms.

Média	Desvio Padrão	Mediana	Máximo	Mínimo
50	72	32	1688	0

A *figura 2* traz a distribuição de ocorrências das *lags*, em blocos de 6 milissegundos. Excluímos da contagem, para este gráfico, as *lags* com valores maiores que 264 ms, pois a apresentação delas no gráfico inviabilizaria a visualização do eixo y. Deste gráfico foram excluídas 24 *lags*, distribuídas entre 265 ms e 1688 ms, a maior coletada.

Observando-se o gráfico da *figura 2*, juntamente com a *tabela 2*, pode-se ver que, apesar da máxima encontrada apresentar duração de 1688 ms, as *lags* medidas estão muito abaixo deste valor. A partir do cálculo dos percentis dos dados, constatamos que cerca de 90% dos dados são *lags* com valores abaixo de 110 ms e que ainda contam-se aproximadamente 220 *lags* (cerca de 8% dos dados) com valores superiores a este.

27

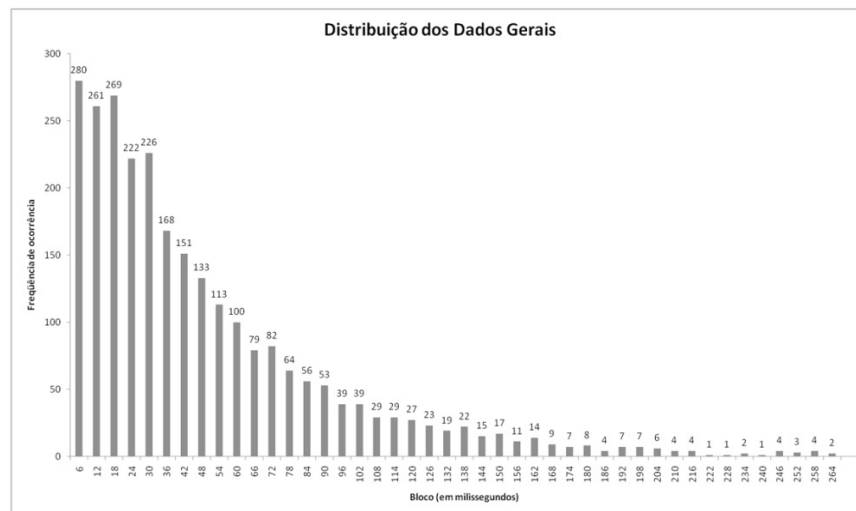


Figura 2: Distribuição dos dados gerais. O eixo x indica intervalos de valores de duração, em ms. No eixo y, assim como nos picos de cada coluna, tem-se a quantidade da ocorrência para cada intervalo de duração.

Obtivemos valores mais altos do que o máximo apresentado. Estes foram excluídos por serem momentos de hesitação ou de “perda da

¹³ O valor zero para duração das *lags* indica sincroniza perfeita. Contamos ao todo 65 *lags* com este valor, ou seja, 65 sincronias perfeitas.

deixa” na realização da tarefa, momentos nos quais não houve interação entre os osciladores envolvidos. Esses foram considerados *outliers* e excluídos dos cálculos estatísticos. Trazemos abaixo um exemplo do sinal acústico de *lag* em torno da média e um de *outlier*:

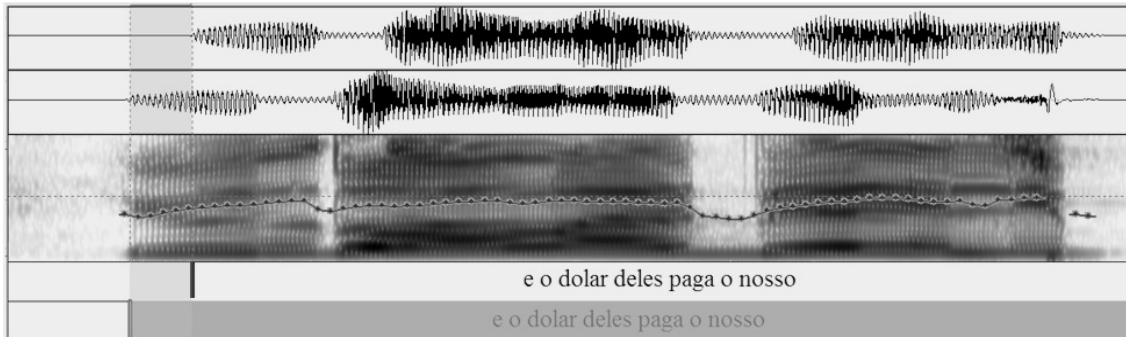


Figura 3: Exemplo de *lag* em torno da média. A seleção no início do sinal indica uma *lag* de 48 ms.

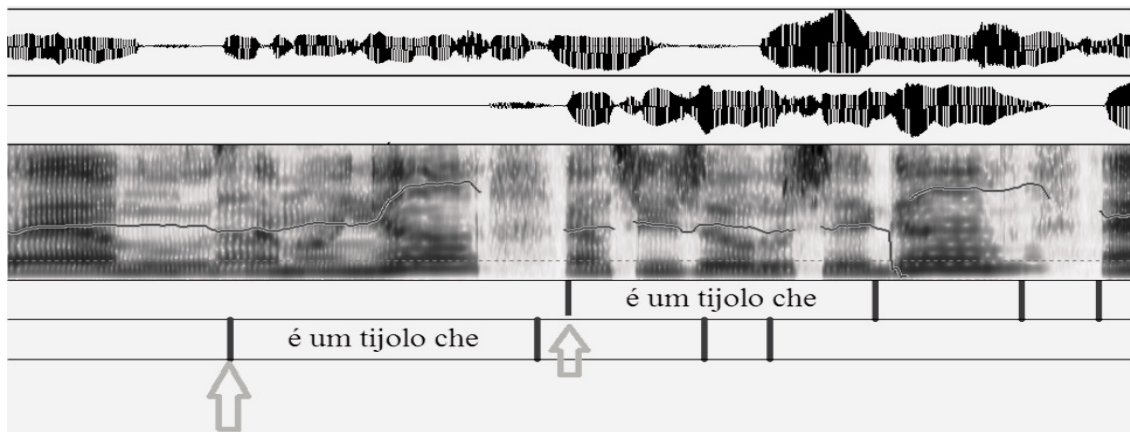


Figura 4: Exemplo de *outlier*. As setas indicam os pontos iniciais. A diferença entre eles é de 1087 ms e representa uma perda de deixa por parte de um dos cantores.

6.1 Resultados de sincronia do fator Ritmo da canção

Em nosso experimento, o fator Ritmo da Canção apresenta dois níveis: Ritmo Sincopado (RS) e Ritmo Não-Sincopado (RNS). A *tabela 3* apresenta a estatística descritiva que resume as medidas das *lags* nesses dois níveis:

Tabela 3: Estatística descritiva do fator Ritmo da Canção. Valores em ms.

	Média	Desvio Padrão	Mediana	Máximo
RS	55	82,5	34	1687,56
RNS	45	59	29,5	1027,18

Pode-se observar que as medidas apresentam valores maiores no nível ritmo sincopado (RS), quando comparadas ao nível rítmico não-sincopado (RNS), muito embora, no caso das médias, em termos absolutos, a diferença de 10 ms seja, a princípio, desprezível. No entanto, estamos ponderando sobre as diferenças encontradas tendo por base que os valores que indicam assincronia são sistematicamente maiores quando se trata do RS.

Os valores de desvio padrão são 82,5 ms e 59 ms para os níveis RS e RNS, respectivamente. Essa diferença entre os níveis indica maior amplitude de variação em torno da média no nível RS em relação ao nível RNS. Em outras palavras, a menor variação na realização da tarefa indica que o processo de sincronização é mais estável no nível RNS, resultado próximo ao que observamos em nosso estudo piloto (Santos & Raposo de Medeiros, 2010). Com tais valores ainda, baseados em suas respectivas médias, chega-se a amplitudes máximas em torno da média 137,5 ms para RS e 104,5 ms para RNS¹⁴.

O gráfico da *figura 5* ilustra essas medidas (média, desvio-padrão e mediana) e a comparação com os valores referentes aos dados gerais:

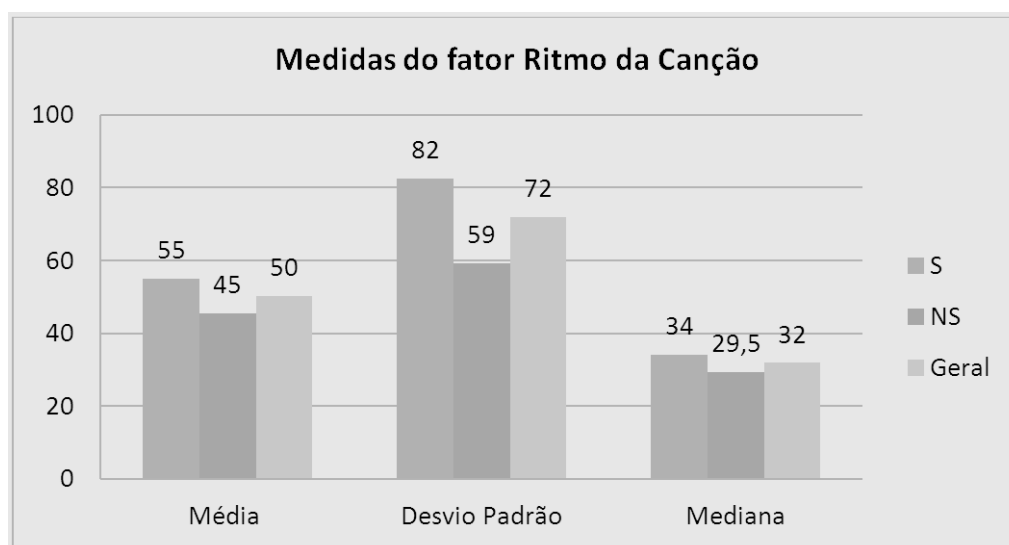


Figura 5: Gráfico descritivo do fator Ritmo da Canção: S representa o nível sincopado, NS o nível não-sincopado. No eixo y tem-se os respectivos valores em ms.

¹⁴ Amplitudes em torno da média a partir do desvio-padrão geralmente são representadas como intervalos. No nosso caso, representaríamos da seguinte forma: [137,54; -27,68] para S e [104,53; -13,79] para NS. Não representamos assim, pois não existem *lags* negativos em nosso experimento, o que pressuporia um líder na realização da tarefa, descaracterizando os cantores como osciladores acoplados, ou seja, apresentamos apenas os valores máximo das amplitudes.

Os gráficos de frequência das *figuras 6 e 7* ilustram a distribuição desses dados:

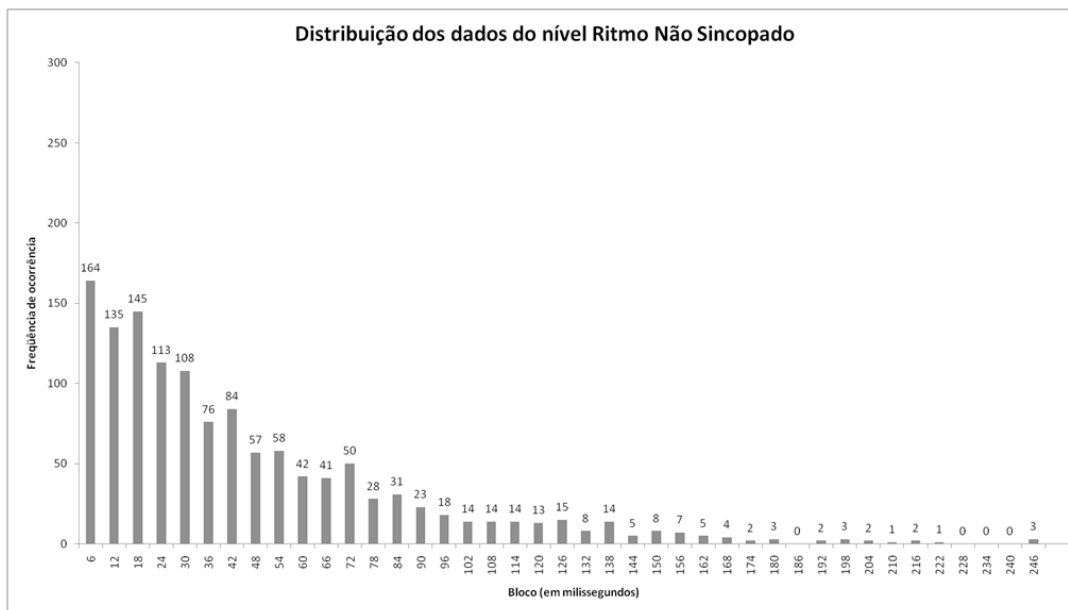


Figura 6: Distribuição dos dados do nível RNS. O eixo x indica intervalos de valores de duração, em ms. No eixo y, assim como nos picos de cada coluna, tem-se a quantidade da ocorrência para cada intervalo de duração.



Figura 7: Distribuição dos dados do nível RS. O eixo x indica intervalos de valores de duração, em ms. No eixo y, assim como nos picos de cada coluna, tem-se a quantidade da ocorrência para cada intervalo de duração

Observando-se os gráficos das *figuras 6 e 7*, nota-se que a distribuição dos dados nos dois níveis, RS e RNS, apresenta distribuição semelhante. Os dados de ambos os níveis apresentam distribuição exponencial com assimetria positiva, ou seja, maior quantidade de dados na região esquerda do eixo x. Uma diferença importante, no

entanto, se apresenta na comparação entre o fator RNS e o fator RS: no fator RNS, as barras mais altas na região esquerda no eixo x , próxima a zero, comparadas às barras menores no fator RS na mesma região, indicam que, no primeiro fator, uma maior quantidade de *lags* (eixo y) apresenta baixo valor de duração (eixo x), reforçando o apresentado pelas medidas estatísticas.

6.2 Resultados de sincronia do fator Ponto de Observação de Sincronização

O fator Ponto de Observação de Sincronização também apresenta dois fatores, denominados Ponto Inicial e Ponto Final. A *Tabela 4* traz a média e o desvio padrão desses níveis:

Tabela 4: Estatística descritiva do fator Ponto de Sincronização

	Média	Desvio Padrão	Mediana	Máximo
PI	63,5	84,7	45	1687,5
PF	36,5	53	24	1027

Observam-se médias com valores de 63,5 ms e 36,5 ms no nível PI e PF, respectivamente, o que vai de acordo com nossa segunda hipótese, já que a média nos pontos iniciais é maior que a dos pontos finais, indicando maior dificuldade de realizar a tarefa de sincronização nos pontos iniciais das sentenças. O desvio-padrão também apresenta valor maior para o nível PI, 63,5 ms contra 36,5 ms para o nível PF, mostrando maior estabilidade de sincronização no nível PF. A partir dos desvios-padrão obtidos, pode-se chegar à amplitude de variação em torno da média de 148,5 ms para PI e 89,5 ms para PF. O gráfico da *figura 8* ilustra a apresentação desses dados:

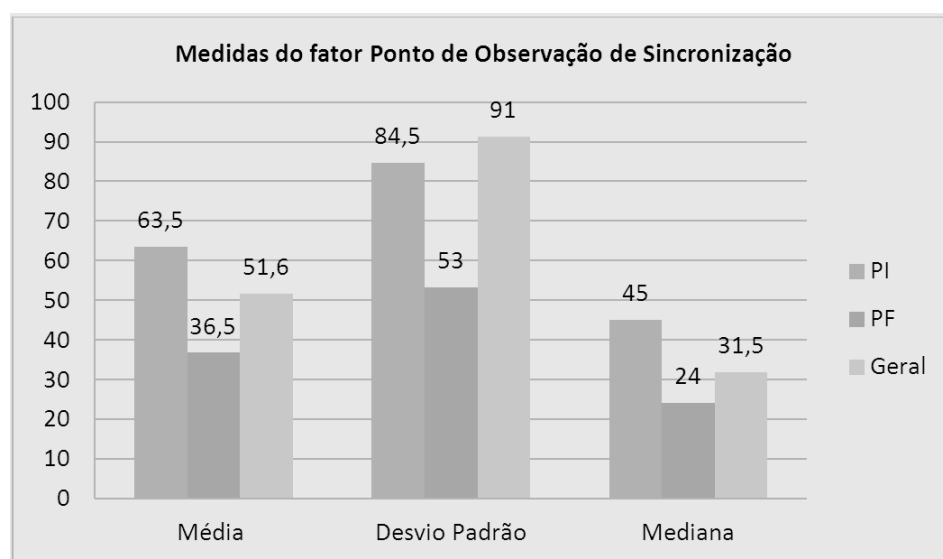


Figura 8: Gráfico descritivo do fator Ponto de Observação de Sincronização

Para ilustrar a distribuição dos dados, trazemos os gráficos de frequências dos dois níveis do fator em questão:



Figura 9: Distribuição dos dados do nível Ponto inicial. O eixo x indica intervalos de valores de duração, em ms. No eixo y, assim como nos picos de cada coluna, tem-se a quantidade da ocorrência para cada intervalo de duração

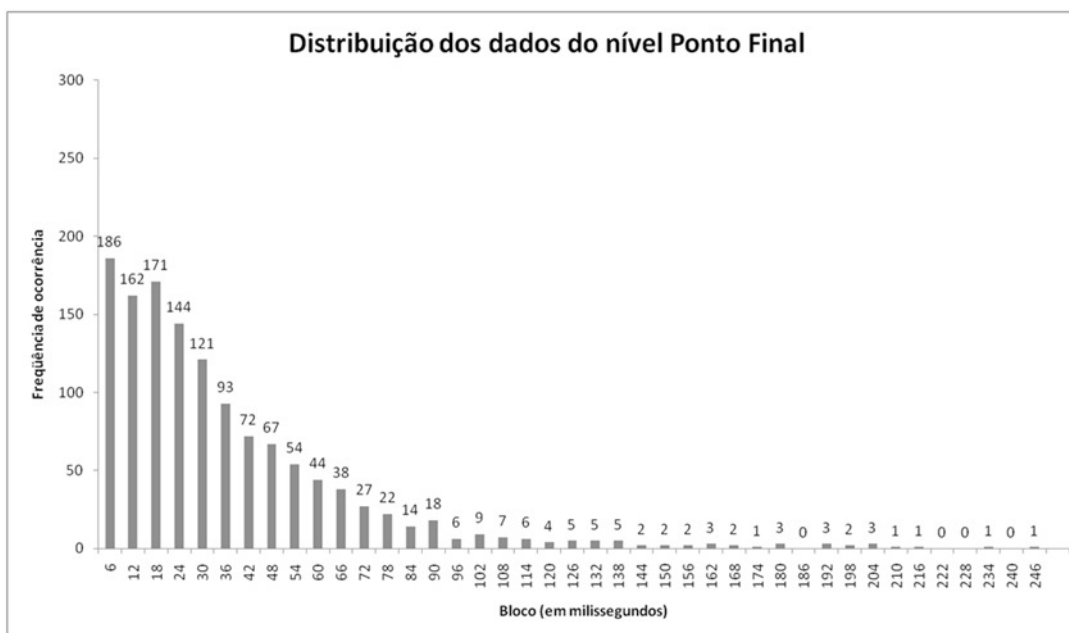


Figura 10: Distribuição dos dados do nível Ponto Final. O eixo x indica intervalos de valores de duração, em ms. No eixo y, assim como nos picos de cada coluna, tem-se a quantidade da ocorrência para cada intervalo de duração.

A mesma comparação feita com os gráficos de distribuição do fator anterior pode ser feita nestes, notando-se, neste fator, que a diferença entre os níveis é maior, o que se ilustra pela maior diferença entre a altura das primeiras barras, próximas a zero.

Como colocamos nas hipóteses, a combinação entre os dois fatores do experimento fez surgir uma terceira hipótese. Para apresentar os resultados das combinações entre os fatores e seus níveis, trazemos a tabela a seguir:

Tabela 5: Combinação entre os níveis dos dois fatores. Valores em ms.

		Ritmo	
		RS	RNS
Ponto	Inicial	69,8	57,3
	Final	40	33,5

Confirmando nossas hipóteses, nota-se que a maior média é a encontrada na combinação dos níveis RS e PI, e que a menor delas é a referente à combinação dos níveis RNS e PF. Esses resultados indicam maior facilidade de sincronização nos pontos finais das canções não sincopadas e maior dificuldade nos pontos iniciais das canções sincopadas.

Contabilizamos, também, as quantidades de sincronias perfeitas, ou seja, de *lags* iguais a zero. A *tabela 6* indica esse cálculo:

Tabela 6: Quantidade de sincronias perfeitas nos dois ritmos musicais e nos dois pontos da sentença.

Total	RS	RNS	PI	PF
65	26	39	28	37

Nota-se que, de acordo com nossas primeira e segunda hipóteses, encontramos maior quantidade de sincronias perfeitas nos níveis RNS e PF, indicando maior facilidade de realização da tarefa nesses níveis, já que pedimos aos sujeitos que cantassem o “mais sincrônico possível”.

Para finalizar a apresentação dos resultados apresentamos o gráfico da *figura 11*, que, ilustrando a consistência do fenômeno investigado, apresenta a relação entre as duas observações realizadas a respeito da sincronização no canto: a devida ao tipo de ritmo e a devida ao ponto observado na sentença. Como já mostrado, o ritmo complexo, RS, apresenta os valores maiores de duração das *lags* e o ritmo simples, RNS, apresenta os valores mais baixos. As linhas de tendência mostram que a melhora da sincronia vai na mesma direção para os dois tipos de ritmo. Se as linhas se cruzassem, teríamos que explicar uma interação entre tipo de ritmo e ponto observado na sentença. Mas isso não ocorreu.

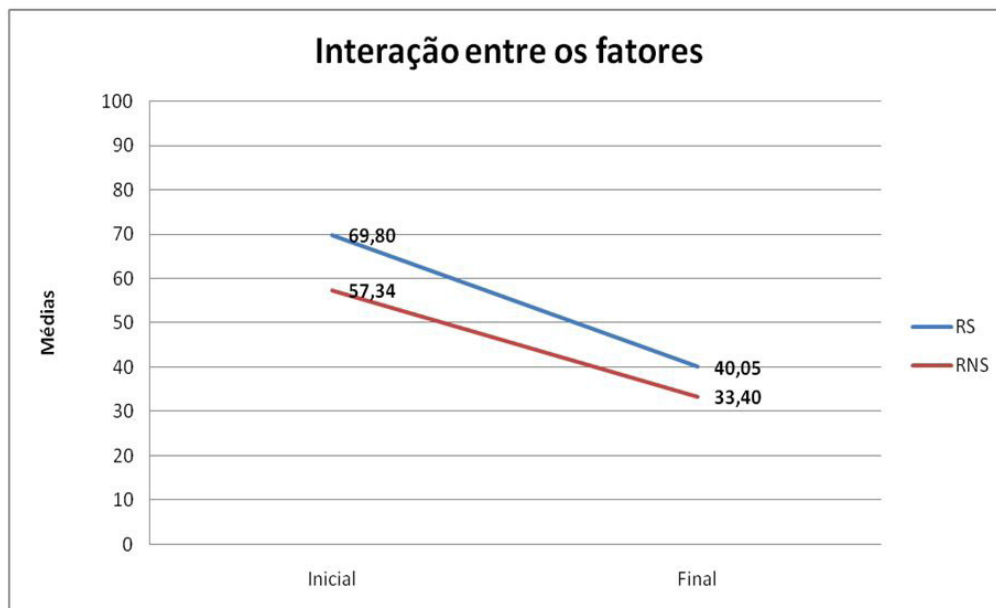


Figura 11: Gráfico de interação entre fatores: apresenta duas linhas de tendência, uma para cada tipo de ritmo. Valores em ms.

7. Discussão

Os dados obtidos não apresentam distribuição normal ou distribuição *T* de *Student*, e por esta razão fomos parcimoniosos em relação a submetê-los a um teste de hipóteses neste estudo preliminar. A distribuição aqui observada, exponencial e com assimetria positiva, deve-se ao fato de não haver um líder na tarefa de sincronização¹⁵, proporcionando-nos apenas resultados positivos nas medidas das *lags*.

Outra consequência desta distribuição é o fato de as diferenças entre as médias parecerem não indicar uma real diferença entre os níveis, o que se deve, também, ao fato de as *lags* terem passado por um cálculo para igualar as proporções dos andamentos das gravações. Nestes cálculos, padronizamos os valores dos andamentos em 120 batidas por minutos, valor mais rápido quando comparado à fala falada e que influencia diretamente os valores das *lags*, já que, quando maior o andamento, menor o valor das *lags*.

Para entender as diferenças rítmicas entre o *rock* e a bossa-nova, observamos, primeiramente, a estrutura musical das canções como um todo, em busca daquilo que, para influenciar de igual maneira a realização da tarefa nos quatro contextos experimentais—os níveis dos fatores—, estivesse presente (i) do início ao fim das canções, (ii) de forma contínua, (iii) em momentos de canto ou de pausa.

¹⁵ Se tivéssemos como ponto de vista o fato de que um dos sujeitos de cada dupla fosse o líder na tarefa (o que iria contra a ideia de osciladores acoplados, e por isso não fizemos deste modo) a medição das *lags* também partiria deste princípio, e assim, caso o segundo sujeito da dupla cantasse atrasado em relação ao líder, teríamos *lags* negativas e, caso se adiantasse, teríamos *lags* positivas.

Buscamos ver se a estrutura rítmica musical guiava da mesma forma a sincronização nos inícios e nos fins das sentenças. Hipotetizamos que os inícios de sentenças seriam mais difíceis de serem sincronizados quando comparados aos finais das mesmas, hipótese esta inspirada nos estudos de Fred Cummins a respeito da fala falada. Nossa amostra nos permite concluir que realmente há diferenças de sincronização entre os pontos iniciais e os pontos finais.

Acreditamos que as estruturas temporais da música e da fala atuem em uma espécie de negociação na realização da tarefa de sincronização do canto. Podemos, assim, dizer que nos pontos iniciais houve menor sincronização devido à pausa musical entre uma sentença e outra, o que obrigava os sujeitos a iniciar um novo processo de sincronização a cada sentença, pois o silêncio pode romper o elo entre os cantores. Isso talvez não acontecesse na fala, em que, embora haja pausas, não são pré-definidas e podem mesmo não ocorrer entre sentenças.

8. Conclusão

Assim, observando os resultados do experimento, concluímos, de modo preliminar, que o ritmo do *rock* permite maior facilidade na tarefa de sincronização da fala cantada do que o ritmo da bossa-nova. Todo o resumo estatístico corrobora nossas hipóteses, ou seja, média, desvio-padrão e mediana apresentam valores mais elevados no nível RS do que no nível RNS, assim como também se observa com o nível PI em relação ao PF.

É necessário explicitar a natureza preliminar deste estudo. Primeiramente, as ideias de ritmo sincopado e não sincopado mantiveram-se para fazer jus ao arrazoado de Santos (2012), que leva em consideração o ritmo básico de um e de outro ritmo musical em foco neste trabalho, a bossa nova de um lado e o *rock* de outro. Isto não quer dizer que não há síncopas no *rock*. Neste sentido é melhor, então, pensar em termos de ritmo simples e ritmo complexo, quando se fala da interação canto e fala com foco na sincronização (Raposo de Medeiros, 2014; Raposo de Medeiros & Cummins, 2014). Em segundo lugar, o número de canções deveria ser maior, assim como o número de pares de sujeitos, a fim de se rodar um teste de hipóteses estatístico, caracterizando o estudo como definitivo. Ainda assim, o estudo foi cuidadoso experimentalmente e levantou uma questão de organização temporal importante na via interdisciplinar em que a fala reside, ou seja, a fala na canção, ou como nomeamos, a fala cantada.

Em suma, o estudo aponta que as assincronias observadas revelam que, na tarefa de cantar junto, o conflito entre as estruturas métrica e rítmica acaba por dificultar a realização da tarefa de sincronização. Assim, retomando a ideia de negociação ou *trade-off* estabelecidos na interação fala-canto, podemos dizer que na sincronização de cantores predomina a estrutura rítmica musical e não a da fala.

Por um lado, estudos linguísticos atestam que a fala não seja isócrona e que tenha estrutura rítmica dificilmente classificável, como vimos na seção 3, e que mesmo assim sincroniza com facilidade. Por outro lado, este estudo aponta que a fala cantada comporta-se diferentemente, pois é influenciada pelos ritmos pré-determinados da música, o que fica também evidenciado em Raposo de Medeiros e Cummins (2014).

Referências

- Abercrombie, D. (1967). *Elementals of general phonetics*. Edimburgh: Edinburgh University Press.
- Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala*. Campinas: Pontes Editores; FAPESP.
- Calegare, A. J. A. (2001). *Introdução ao delineamento de experimentos*. São Paulo: Edgard Blücher.
- Clayton, M., Sager, R., & Will, U. (2004). *In time with the music: The concept of entrainment and its significance for ethnomusicology*. ESEM CounterPoint, Vol. 1.
- Cummins, Fred. (2002). On synchronous speech. *Acoustic Research Letters Online*, 3(1), 7–11.
- Cummins, Fred. (2003). Practice and performance in speech produced synchronously. *Journal of Phonetics*, 31(2), 139–148.
- Cummins, Fred. (2006). Measuring synchronization among speakers reading together. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics* (pp. 28–30). Atenas, Grécia.
- Cummins, Fred. (2009a). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37(1), 16–28. Elsevier.
- Cummins, Fred. (2009b). Rhythm as an affordance for the entrainment of movement. *Phonetica*, 66(1–2), 15–28.
- Dauer, R. M. (1983). Stress-timing end syllable-timing reanalyzed. *Journal of Phonetics* 11, 51–62.



- Lieberman, M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Lieberman, M., & Prince, A. (1977). On Stress and linguistic rhythm. *Linguistic Inquiry*. Vol. 8, N. 2, 249–336.
- Major, R. C. (1981). Stress-timing in Brazilian Portuguese. *Journal of Phonetics*, v. 9 n° 3, 343–352.
- Mayberry, R. I., Jaques, J., & Dede, G. (1998). What stuttering reveals about the development of the gesture-speech relationship. In J. M. Iverson and S. Goldin-Meadow (Orgs.), *The nature and function of gestures in children's communication*. New Directions for Child Development, n. 79. San Francisco: Jossey-Bass.
- Pike, K. (1971). *Phonemics: A technique for reducing languages to writing*. Ann Arbor: The University of Michigan Press.
- Port, R. F., Tajima, K., & Cummins, F. (1996). Self-Entrainment in animal behavior and human speech. *Online Proceedings of the 1996 Midwest Artificial Intelligence and Cognitive Science Conference*.
- Raposo de Medeiros, B. (2002). *Descrição comparativa de aspectos fonético-acústicos selecionados da fala e do canto em português brasileiro*. (Tese de Doutorado). Campinas: UNICAMP.
- Raposo de Medeiros, B., & Cummins, F. (2014). Speech and song synchronization: A comparative study. In *Proceedings of the Seventh International Conference on Speech Prosody*. Science Foundation Ireland. (pp. 748–751).
- Raposo de Medeiros, B. (2014). A organização temporal na fala e no canto: Estudo com foco na sincronização. *Percepta* 1(2), 27–50.
- Santos, C. A. A. & Medeiros, B. R. (2010). Interações entre ritmo linguístico e ritmo musical na canção. *Anais do VI SIMCAM, Simpósio de Cognição e Artes Musicais* (pp. 279–291). Rio de Janeiro.
- Santos, C. A. A. (2012). *A sincronização em dois ritmos da canção: Uma observação experimental acerca da fala cantada*. (Dissertação de Mestrado.) Faculdade de Filosofia, Letras e Ciências Humanas. São Paulo: Universidade de São Paulo.
- Schönberg, A. *Fundamentos da composição musical*. São Paulo : Edusp, 1996.