

## 1) Assincronia vs. módulo da assincronia

Se vocês acompanharam os últimos capítulos da análise, perceberam que eu sempre utilizei o módulo da assincronia como variável independente nos modelos estatísticos. Quando Mauricio e eu mostramos a Barbara Tillmann os resultados preliminares, ela logo perguntou : "porque o módulo"? A minha resposta foi que a distribuição da verdadeira assincronia era simétrica e com média e mediana próximas de zero (cf gráfico em um dos meus mails anteriores).

Mas a Barbara tem razão pois, ao usarmos o módulo da assincronia, nós eliminamos a possibilidade de anlaisar os adiantamentos e atrasos de sincronismo. Por exemplo, pode ser que os clarinetistas adiantem sistematicamente quando na condição "self" e atrasem na condição "other".

Para tirar todas as dúvidas, eu fiz a análise estatística com o valor real da assincronia. Como de hábito, o script em R está anexado a este. Os efeitos fixos self/other e take não são significativos, entretanto, os efeitos aleatórios primo e secondo são significativos. Isto significa que alguns sujeitos atrasam ou adiantam sistematicamente (efeito secondo) e que alguns sujeitos primos fazem com que os segundos que os acompanham atrasem ou adiantem sistematicamente (efeito primo). Os valores obtidos para cada clarinetista estão no gráfico em anexo. Valores positivos significa atraso e valores negativos significa avanço. Tirando o Walter, existe uma correlação negativa entre os dois efeitos que poderia ter a explicação seguinte: quando um clarinetista tem a tendência a atrasar a sincronia quando toca secondo, ele tem a tendência de fazer adiantar os outros quando ele toca primo.

## 2) Transformação dos dados

A distribuição de probabilidade dos valores de assincronia (em módulo) medidos é altamente assimétrica (cf gráfico que enviei em um dos meus mails anteriores). Para que o GLMM funcione corretamente, os dados precisam ser transformados de maneira que a distribuição seja o mais aproximadamente gaussiana o possível. Existe uma técnica muito simples para tanto chamada "Box & Cox power transform":

[http://en.wikipedia.org/wiki/Power\\_transform](http://en.wikipedia.org/wiki/Power_transform)

Simplificando um pouco, esta transformação realiza a transformação monotônica e inversível  $y = x ^ \lambda$ . R tem uma função para achar o valor ótimo de lambda que fará a distribuição da variável transformada y se tornar o mais normal possível. O segundo arquivo com código R anexado a este ilustra esta transformação. O valor obtido para lambda está por volta de 0.3. Os gráficos anexados mostram as distribuições antes e após a transformação de Box & Cox. Os gráficos anexados mostram as distribuições antes e após a transformação de Box & Cox.

A distribuição não passa pelo teste de normalidade de Shapiro-Wilks, mas é bem menos assimétrica quando transformada, com um coeficiente de skeweness por ordem de 20 vezes menor.

De qualquer maneira, o efeito claro da transformação é o fato das amostras com alto valor de assincronia (algumas acima de 150 ms, o que são certamente erros de execução) passarem a ter menos importância no

ajustamento do modelo. Isto evita os inconvenientes de a um procedimento de rejeição de casos extremos (outliers) ou de regressão robusta.

Uma nota final: em todo rigor, o que é exigido para a GLMM é que os \*\_resíduos\_\* do ajustamento do modelo sejam distirbuídos de maneira normal. O que eu descrevo acima é uma simplificação.

Mais capítulos seguirão.

Abs,

Rafael

=====

Hani Camille Yehia <[hani@cefala.org](mailto:hani@cefala.org)> [2011-01-10 15:55]:

> Apenas a título de querer dar um "pitaco" no meio da discussão, a  
> figura enviada pelo Rafael (e reproduzida ao final desta mensagem)  
> parece indicar que a assincronia com "self" tende a diminuir com o  
> tempo, em contraste com o "other" onde, pelo menos, parece não haver  
> aprendizado.

Eu já tinha observado isto, mas não sei como explicar o fenômeno. Note que o excerto de Stravinsky dura apenas uns 15 segundos. Entretanto, note também que no experimento os clarinetistas tocam quatro vezes (os tais "takes"). Na análise estatística que eu fiz, obtive um efeito significativo para o fator take, a sincronia global melhorando de 2 ms a cada take. Por fim, note também que o gráfico que eu enviei mostra as médias para todos os takes.

De qualquer maneira, quando eu refizer as análises, vou também introduzir um fator contínuo "tempo" (no interior de cada take) para testar a sua hipótese.

Este experimento abre mais portas do que ele fecha. Os resultados dão espaço para várias interpretações. Por exemplo, o efeito self/other que observamos pode ser devido unicamente a uma memória motora, pois o intervalo entre a execução dos primos e dos segundos foi de apenas um dia. É por isto que Keller et al. (2007) fizeram as gravações de primo e segundo espaçadas de 2 a 3 meses.

A única maneira de avançar nestas questões será através de novos experimentos. É trabalhoso, mas isto é que é legal na ciência.

Voltando ao assunto dans notas no primeiro tempo: eu me enganei totalmente no particionamento das notas. Acredito que as notas no primeiro tempo dos compassos sejam: 11, 15, 25 e 41. Tem também um primeiro tempo antes da nota 32, que cai no contratempo. Segue abaixo o gráfico corrigido.

Abraço,

Rafael

```
@Article{Keller_et_al._2007,
  author = "Keller, Peter E. and Knoblich, Günther and Repp,
            Bruno H.",
  title = "Pianists duet better when they play with themselves:
            on the possible role of action simulation in
            synchronization",
  journal = {Conscious Cogn},
  year = 2007,
  volume = 16,
  number = 1,
  pages = "102-111",
  PMID = 16466932
}
```

=====

Em 10-1-2011

Conversei com o Maurício hoje de manhã e discutimos sobre os resultados que enviei ontem. A questão da escolha de repartição das notas está ainda em suspenso. Eu escolhi as notas de início de compasso mas isto ainda deve ser discutido entre nós.

Só para dar mais subsídios para a nossa reflexão, segue abaixo um gráfico com os valores de assincronia para cada nota. Como de hábito, o script em R para gerar a figura também está anexado a este.

Nesta figura, cada ponto corresponde à assincronia média para todos segundos e todos takes, separadas em "self" (azul) e "other" (vermelho). As barras verticais indicam o erro padrão. As linhas verticais cinzas indicam as notas de início de cada compasso.

Abraço,

=====

Respondendo com cópia ao Maurício e ao Hani  
Em 9-1-2011

Davi,

Obrigado pela tabela corrigida. Ontem eu andei brincando com os dados fazendo a mesma correção de lag que você fez agora. Já fiz várias análises preliminares em R, mas vou retomá-las nesta semana. Pelo momento, estou enviando em anexo a figura da distribuição da assincronia nos onsets, obtida com o código em R abaixo:

```
onset.data <- read.table ("raw_data", header = FALSE) names (onset.data)
<- c ("primo", "secondo", "composer", "take", "note", "primo.onset",
      "secondo.onset") async <- onset.data$secondo.onset -
onset.data$primo.onset plot (density (async / 44100, na.rm = TRUE), las =
1, main = "", xlab = "onset asynchrony (s)") abline (v = 0, col = "gray")
dev.copy2pdf (file = "asynchrony-distribution.pdf")
```

A média das assincronias é ligeiramente inferior a zero (-1.4 ms), o que significaria que os segundos tendem a antecipar os primos. Entretanto, o pico da distribuição é por volta de +5.6 ms, o que significaria atraso na execução. Esta diferença entre a média e o pico da distribuição acontece porque a distribuição é altamente assimétrica ("skewed", em inglês) e está longe de ser uma distribuição gaussiana, ou normal.

De qualquer maneira, isto não deverá ser problemático para a análise. Isto também levanta um aspecto muito importante da análise: acho que faria mais sentido considerar o módulo da assincronia, cujo valor seria sempre positivo. Acho que foi isto que vocês fizeram quando analisaram os IOIs, não é mesmo?

De qualquer maneira, vocês encontrarão anexados a este mais dois gráficos que mostram o efeito "ego/alia" ("self/other" em inglês; como se diria em português?), o primeiro para a assincronia em valor natural e o segundo para o módulo da assincronia. Estes gráficos foram obtidos com o código seguinte em R:

```
who <- rep("other", nrow(onset.data))
idx <- which(onset.data$primo == onset.data$segundo) who[idx] <- "self"
onset.data$who <- factor(who)
onset.data$asynchrony <- async
boxplot(asynchrony / 44100 ~ who, data = onset.data, ylab = "asynchrony (s)", las = 1) dev.copy2pdf(file = "asynchrony-who.pdf", width = 3, height = 7)
boxplot(abs(asynchrony) / 44100 ~ who, data = onset.data, ylab = "|asynchrony| (s)", las = 1) dev.copy2pdf(file = "abs-asynchrony-who.pdf", width = 3, height = 7)
```

Note que as médias entre self e other para o valor natural da assincronia não são muito diferentes de zero nem muito diferente entre elas. Entretanto, a variabilidade é maior para o caso other e é isto que conta. O segundo gráfico mostra as distribuições do módulo da assincronia para as condições self e other.

Bom, eu já tinha feito as análises estatísticas ontem com os dados corrigidos em lag, mas de maneira aproximada. Vou refazer tudo com a tabela que você enviou hoje. Só para quebrar um pouco o suspense, aqui vão alguns resultados que já obtive, utilizando o valor em módulo da assincronia:

- 1) O efeito self/other é significativo. O módulo da assincronia é por volta de 10ms menor na condição self.
- 2) O efeito take é significativo. Em média, os clarinetistas diminuem o módulo da assincronia de uns 2ms a cada take, o que significa que os segundos se adaptam aos primos, o que é um bom sinal.
- 3) Eu classifiquei as notas entre aquelas que caem no primeiro tempo de cada compasso (notas 5, 11, 15, 22, 31 e 39, se não me enganei, por favor verifique) e as outras. O efeito também é significativo, a assincronia diminui de 5ms para as notas no primeiro tempo (bom, teremos ainda que discutir sobre isto; eu acredito que a minha maneira de classificar as notas é pertinente, mas eu não sou músico...)
- 4) Um outro resultado periférico, mas altamente interessante: o fator

"primo", considerado como um fator aleatório e não como um fator fixo (explicarei a diferença numa outra oportunidade) é também significativo. Isto quer dizer que alguns primos são mais fáceis de acompanhar que outros. O primo "mais confortável" é o Dominique, que faz a assincronia média baixar de 4.6 ms. O primo "menos confortável" é o Walter, que faz a assincronia aumentar de 3.9 ms, em média.

- 5) Na análise que estou fazendo com GLMM ("general linear mixed model"), que é superior à análise ANOVA clássica, dá também para testar o efeito aleatório (e não fixo) do "secondo". O craque no acompanhamento é o Marcus, cuja assincronia é 13 ms abaixo da média. O pior acompanhador está a 13 ms acima da média, mas não vou dizer quem é...

Bom por hora é só. Com o que temos como resultados por hora, já dá para pensar em publicar um full paper.

Abraço,

Rafael

\* Davi Mota <[davialvesmota@yahoo.com.br](mailto:davialvesmota@yahoo.com.br)> [2011-01-09 03:02]:

> Assim ficou fácil de resolver, já tava ficando preocupado.

>

> Até mais

>

> Davi

>

>

>

>

> De: Rafael Laboissiere <[rafael@laboissiere.net](mailto:rafael@laboissiere.net)>

> Para: Davi Mota <[davialvesmota@yahoo.com.br](mailto:davialvesmota@yahoo.com.br)>

> Enviadas: Sábado, 8 de Janeiro de 2011 19:14:33

> Assunto: Re: Problemas de sincronismo

>

> Davi,

>

> Já descobri de onde vem o problema. Nas gravações de áudio de  
> qualidade, o secondo está tipicamente 210 ms adiantado com relação ao  
> primo. Sendo assim, você deve subtrair os valores de lag dos valores  
> medidos de onset de maneira a obter a coluna onset\_secondo.

>

> A culpa é minha, pois eu acho que eu tinha te dito há alguns dias para  
> adicionar os valores, ao invés de subtrair.

>

> Abraço,

>

> Rafael

>

> \* Rafael Laboissiere <[rafael@laboissiere.net](mailto:rafael@laboissiere.net)> [2011-01-08 21:05]:

>

> \* Davi Mota <[davialvesmota@yahoo.com.br](mailto:davialvesmota@yahoo.com.br)> [2011-01-08 10:16]:

>

>

```

> > > segue o anexo com os valores de onsets, com as correções discutidas
> > > O formato é:
> > >
> > >      primo secondo compositor take onset_primo onset_segundo
> > >
> > > Acredito que ainda encontraremos alguns problemas, como valores de
> > > assincronia muito elevados, resultante de erros de execução.
> > > Talvez precisemos escolher algumas notas chave como o Keller fez
> > > para ter resultados mais consistentes.
> >
> > O código R abaixo produz a figura anexada com o histograma de das
> > asincronias (onset_segundo - onset_primo).:
> >
> > d <- read.table ("raw_data", header = FALSE) names (d) <- c
> > ("primo", "segundo", "composer", "take", "note", "primo.onset",
> > "segundo.onset")
> > plot (density ((d$segundo.onset - d$primo.onset) / 44100, na.rm =
> > TRUE), main =
> > "")
> >
> > A distribuição tem um pico por volta de 0.4 s e um segundo pico
> > menor por volta de 0.8 s. Acredito que algo esteja errado com os
> > dados. Este valor de 0.4 s é aproximadamente o dobro do valor típico
> > de lag que eu encontrei (9300 amostras, que corresponde a 0.21 s).
> > Talvez você tenha adicionado o lag na coluna errada.
> >
> > Até +
> >
> > Até
>
>
>
> Rafael
>
>
>
>

```

Rafael

-----

The raw values of the absolute asynchrony are represented in the boxplots of figure [1]. A GLMM (generalized linear mixed model) was fitted to the data. The fixed factors of this model were "who" (a discrete factor with levels "self" and "other") and "take" (a continuous factor going from 1 to 4). The fitted values for the fixed effects are represented with filled circles (blue for "self" and red for "other"). The random factors are the identity of the primo performer and the identity of the segundo performer.

A zoomed plot for the fitted values is shown in figure [2], together with the standard error around the estimated values. The difference between the self and other asynchronies for take #1 is of 9.9 ms. This value is significantly different from zero (95% confidence interval: [6.2, 13.7],  $p < 0.0001$ ). For the "other" condition, there is a significant change of the asynchrony with the take, with a decrease of 1.4 ms per take (95% confidence interval: [0.6, 2.3],  $p < 0.01$ ). The change of asynchrony with take is not significant for the "self" condition, although there is a slight decrease of 0.4 ms per take.

The significance of the random effects were tested using a likelihood ratio test. Both the primo (Chisq [1] = 31.4,  $p < 0.0001$ ) and the secondo (Chisq [1] = 258.5,  $p < 0.0001$ ) random factors are significant. The fitted values for each of the six performs are shown in figure [3]. The horizontal axis show the primo effect and the vertical axis show the secondo effect. The individual values of the primo and second random effects have the following interpretation: a performer that has a positive value for the primo effect induces a increase of asynchrony on the other performers, i.e. it is difficult to follow him (or her). The performer who has a negative value for the secondo random effect is one who can better follow the primo.