

Davi Alves Mota

***Análise dos padrões de sincronização em duos
de clarineta a partir de parâmetros acústicos
e cinemáticos***

Belo Horizonte

2012

Davi Alves Mota

***Análise dos padrões de sincronização em duos
de clarineta a partir de parâmetros acústicos
e cinemáticos***

Dissertação apresentada ao Programa de
Pós-Graduação da Escola de Música da
Universidade Federal de Minas Gerais
como requisito parcial para a obtenção
do título de Mestre em Música.

Linha de Pesquisa: Sonologia

Orientadores:

Maurício Loureiro

Rafael Laboissière

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE MÚSICA

Belo Horizonte

2012

M917a Mota, Davi Alves.

Análise dos padrões de sincronização em duos de clarineta a partir de parâmetros acústicos e cinemáticos [manuscrito] / Davi Alves Mota. – 2012.
64 f., enc.

Orientadores: Maurício Loureiro e Rafael Laboissière.

Dissertação (mestrado em Música) – Universidade Federal de Minas Gerais, Escola de Música.

Inclui bibliografia: p. 61-64.

1. Música para clarinete. 2. Sincronização (música). 3. Acústica (música). 4. Teoria musical. I. Loureiro, Maurício. II. Laboissière, Rafael. III. Universidade Federal de Minas Gerais. Escola de Música. IV. Título.

CDD: 785.701

Resumo

Este estudo investiga as estratégias de sincronização utilizadas na performance instrumental de conjuntos musicais, partindo de duas abordagens complementares, análise dos movimentos corporais e análise de parâmetros acústicos. Duas hipóteses sustentam este estudo. A primeira prevê que os músicos atinjam um melhor desempenho em termos de sincronização dos parâmetros musicais quando eles acompanham a si mesmos (efeito *self-other*). A segunda prevê a existência de um acoplamento entre a movimentação dos músicos e o resultado acústico produzido durante a performance. Uma combinação destas hipóteses prevê que o efeito *self-other* anteriormente observado com dados acústicos poderia igualmente ser detectado nos dados gestuais, mesmo que não haja contato visual entre os músicos durante a performance. Isso seria possível se o clarinetista no papel de "seguidor" adaptasse de alguma forma o seu padrão gestual ao padrão gestual do clarinetista "líder".

As seguintes questões são propostas: (1) A sincronização de parâmetros acústicos se reflete na movimentação dos músicos? (2) Esta sincronização também se reflete em outras dimensões acústicas? Para tanto, foram analisadas performances de clarinetistas profissionais em um ambiente orquestral simulado. Foi realizado um experimento dividido em duas sessões: Na primeira, os músicos foram instruídos a tocar como primeiro clarinetista (*primos*), isto é, seguindo suas próprias intenções interpretativas. Na segunda sessão, os músicos foram instruídos a tocar como segundo clarinetista (*secondos*), acompanhando as gravações de clarinetistas *primos* da primeira sessão, incluindo as executadas por eles mesmos.

A análise dos dados acústicos foi realizada em cinco parâmetros, são eles: *Índice de Articulação*, *IOI Normalizado*, *Índice de Legato*, *Log do Tempo de Ataque* e *Loudness*. Os resultados mostram que os participantes do experimento tendem a seguir com maior eficiência suas próprias execuções, não só em termos rítmicos, mas também em outras dimensões musicais. Para a análise dos dados gestuais foi definido um parâmetro escalar a partir das posições tridimensionais da campana do instrumento, calculado como sua velocidade instantânea. Os resultados mostram que quando os clarinetistas seguem a si próprios eles tendem a conservar o seu perfil de movimentação "original", assim como gravado nas execuções *solo*. Entretanto, quando os clarinetistas seguem um outro músico, eles tendem a se afastar deste perfil "original". Nestes casos, foi observado que as curvas das execuções dos clarinetistas *secondos* tendem a se adaptar ligeiramente às curvas das execuções dos clarinetistas *primos*.

Abstract

In this study we investigate the synchronization strategies used in performances of instrumental music ensembles, using two complementary approaches, analysis of body movements and analysis of acoustic parameters. This study is supported by two hypotheses: in the first one it is assumed that musicians achieve better performance in terms of synchronization of the musical parameters if they follow themselves (effect known as self-other). The second hypothesis, we assume the existence of some sort of coupling between the musician's body movements and the acoustic results produced during performance. The combination of these hypotheses could indicate that the self-other effect, previously observed using acoustic data, could also be detected in musicians' body movements, even without any visual contact during the performance. This would be possible if the player assuming the role of "follower" could somehow adapt his gestural pattern to the gestural pattern of the "leader" clarinetist.

This study attempts to answer the following questions: (1) Is the synchronization of acoustic parameters reflected in the movement of musicians? (2) Is this synchronization also reflected in other acoustic features? To answer this questions, we analyzed the performances of professional clarinetists in a simulated orchestral environment. We conducted one experiment divided in two sessions: In the first session, musicians were instructed to play as first clarinetist (*primo*), i.e., following their own interpretative intentions. In the second session, musicians were instructed to play as second clarinetist (*secondos*), following the recordings of *primo* clarinetists from the first section, including those executed by themselves.

Analysis of acoustic variables was performed on five parameters: *Articulation Index*, *Normalized IOI*, *Legato Index*, *Log Attack Time* and *Loudness*. The results show that the participants tend to follow more efficiently their own recordings, not only in rhythmical terms, but also in other musical dimensions. For gesture data analysis, a single scalar parameter was defined using the three-dimensional positions of the clarinet bell, calculated as the instantaneous speed. The results indicate that when clarinetists follow themselves they tend to retain their "original" gestural profile, as recorded in the *solo* executions. However, when they follow other clarinetists, they tend to deviate from this "original" profile. In these cases, it was observed that the curves of *secondos* clarinetists tend to slightly adapt to the curves of *primo* clarinetists.

Sumário

Resumo	i
Abstract	ii
Sumário	iii
Lista de Figuras	v
Lista de Tabelas	vii
1 Introdução	1
1.1 Gesto e música	2
1.2 Definição de gestos acompanhantes	4
1.3 Objetivos	4
1.3.1 Justificativa	5
2 Materiais e Métodos	6
2.1 Escolha do instrumento	6
2.2 Participantes	7
2.3 Material musical	7
2.4 Experimento	7
2.5 Aquisição e processamento dos dados gestuais	8
2.5.1 Parametrização	9
2.5.2 Definição da região de análise	9
2.5.3 Suavização dos <i>perfis de velocidade</i>	10
2.5.4 Normalização da amplitude	10
2.5.5 Ajustamento Temporal	11
2.6 Aquisição e processamento dos dados acústicos	13
2.6.1 Gravação	13
2.6.2 Segmentação	14
2.6.3 Parametrização do áudio	14
2.6.4 Exclusão de valores inválidos	19

2.7	Análise dos dados	19
2.7.1	Análise Discriminante Linear	20
2.7.2	Redução da Dimensionalidade: Análise em Componentes Principais	21
2.7.3	Retro-Projeção	22
3	Resultados	24
3.1	Discriminação dos dados gestuais	24
3.1.1	Agrupamento por <i>secondos</i>	24
3.1.2	Agrupamento por <i>primos</i>	32
3.2	Discriminação dos dados acústicos	35
3.2.1	Considerações iniciais	35
3.2.2	Discriminação	35
3.3	Determinação de regiões significantes por retro-projeção	38
3.4	Efeito <i>self-other</i>	41
3.4.1	Dados acústicos	41
3.4.2	Dados cinemáticos	43
3.5	Avaliação da discriminação de performances	47
3.5.1	Agrupamento por <i>SECONDOS</i>	48
3.5.2	Agrupamento por <i>PRIMOS</i>	51
4	Discussão	54
5	Conclusão e perspectivas futuras	59
	Referências Bibliográficas	61

Lista de Figuras

2.1	Excerto da “Dança do Camponês e o Urso”	7
2.2	Superposição de dois <i>perfis de velocidade</i> executados com diferentes andamentos.	11
2.3	Ajustamento temporal de <i>perfis de velocidade</i> executados com diferentes andamentos.	12
2.4	Superposição de 24 <i>perfis de velocidade</i> executados pelo mesmo músico, antes e depois do ajustamento temporal.	13
2.5	Representação do cálculo do <i>Índice de Articulação</i> para a transição entre duas notas.	16
2.6	Cálculo do <i>Índice de Legato</i> para a transição entre duas duas notas.	17
2.7	Representação ilustrativa do uso combinado PCA-LDA.	22
3.1	Resultados do LDA para o agrupamento por <i>secondos</i>	25
3.2	Retro-projeção do componente LD2 calculado com o espaço dimensional integral.	26
3.3	Número de classificações erradas para a análise combinada PCA-LDA para o agrupamento por <i>secondos</i>	26
3.4	Resultados da LDA calculada para o agrupamento por <i>secondos</i> a partir dos dados transformados pela PCA.	27
3.5	Contribuição de cada componente LD para cada clarinetista <i>secondo</i>	28
3.6	Superposição dos <i>perfis de velocidade</i> de todas as execuções gravadas pelo clarinetista C4	29
3.7	Retro-projeção dos componentes LD para o agrupamento por <i>secondos</i> , separadas por cada clarinetista.	31
3.8	Número de erros retornado pela combinação PCA-LDA para o agrupamento por <i>primos</i>	32

3.9	Resultados da LDA para o agrupamento por <i>primos</i> , após redução do espaço dimensional pela PCA.	33
3.10	Retro-projeção dos componentes LD para o agrupamento por <i>primos</i> , separadas por cada clarinetista.	34
3.11	Distribuição dos valores previstos para cada classe nos 5 LDs para o agrupamento por <i>secondos</i>	36
3.12	Distribuição dos valores previstos para cada classe nos 5 LDs para o agrupamento por <i>secondos</i>	37
3.13	Superposição dos valores de significância calculados para os <i>perfis de velocidade</i> e para cinco descritores acústicos.	40
3.14	Boxplot dos descritores acústicos analisados para o efeito <i>self-other</i>	42
3.15	Representação esquemática do procedimento empregado no cálculo das distâncias <i>self-other</i>	45
3.16	Comparação entre as distâncias dos conjuntos de dados $D(S,s)$ e $D(S,o)$ e entre os conjuntos de dados $D(O,s)$ e $D(O,o)$	46

Lista de Tabelas

3.1	Diferenças dos descritores acústicos analisados para o efeito <i>self-other</i> .	43
3.2	Resultado do teste de validação cruzada para a classificação dos dados cinemáticos. Agrupamento por clarinetistas <i>secondos</i>	49
3.3	Resultado do teste de validação cruzada para a classificação dos dados acústicos. Agrupamento por clarinetistas <i>secondos</i>	50
3.4	Resultado do teste de validação cruzada para a classificação dos dados cinemáticos. Agrupamento por clarinetistas <i>primos</i>	51
3.5	Resultado do teste de validação cruzada para a classificação dos dados acústicos. Agrupamento por clarinetistas <i>primos</i>	53

1 Introdução

A manipulação de características temporais em performances musicais é uma estratégia amplamente utilizada pelos músicos para demonstrar suas intenções expressivas. No caso de conjuntos instrumentais, a coordenação dessas manipulações entre os intérpretes é crucial, tanto para a realização da performance quanto para o bom entendimento da ideia musical transmitida ao público. Conseqüentemente, o ônus desta coordenação é compartilhado entre todos os músicos envolvidos na realização da tarefa musical, seja servindo de referência para outros músicos (como é o caso do regente ou chefes de naipes) ou seguindo as diretrizes ditadas por esta referência. Para realizar essa tarefa apropriadamente, os músicos são obrigados, de alguma forma, a antecipar as variações temporais produzidas por outros membros do conjunto.

Durante uma performance em grupo como, por exemplo, em um ambiente de orquestra, os intérpretes tiram vantagem das informações visuais e acústicas, transmitidas de forma ininterrupta pelos outros intérpretes, a fim de melhorar sua sincronização. Por exemplo, no início de uma frase que ocorre após um grande período de silêncio, um flautista poderia indicar aos companheiros de naipe o momento preciso de iniciar, usando para isso um movimento da extremidade da flauta ou alguma indicação com a cabeça. Um caso semelhante ocorre em finais de frase com variações temporais, onde uma falha na coordenação da variação por parte de um dos músicos prejudicaria a percepção do efeito musical esperado.

Pode-se argumentar ainda que, no caso de performances em grupo, as informações visuais possuem papel essencial na sincronização dos eventos musicais. No entanto, músicos também são capazes de acompanhar outros músicos exclusivamente a partir de sinais acústicos. Essa situação de performance se tornou trivial como consequência ao crescimento da indústria fonográfica, onde a demanda por perfeição e necessidade de baixar os custos provocaram o aumento do número

de gravações feitas em estúdio. Por outro lado, diversos tipos de movimentos corporais não diretamente vinculados à produção sonora podem ser constantemente observados em qualquer execução instrumental. De fato, como apontado por Gabriellsson (2003, p. 249), a movimentação realizada por um músico, além de comunicar informações relevantes para a coordenação com outros, pode ainda assumir papéis diversos, como o de comunicar intenções expressivas, apresentar informações sobre a personalidade do artista ou simplesmente entreter a audiência. Nos últimos anos, grandes esforços foram dedicados ao estudo desses movimentos e, apesar de não haver consenso sobre sua origem ou o papel, sua existência é inegável (WANDERLEY; DEPALLE, 2001).

O foco deste estudo é a investigação das estratégias de sincronização utilizadas na performance instrumental de conjuntos musicais. Ao tocar em grupo, o músico deve sincronizar os resultados de suas próprias ações com as ações dos outros. Uma série de estudos recentes demonstraram que músicos acompanham mais facilmente suas próprias gravações (KELLER, 2001; KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012). Em estudos voltados para o reconhecimento da ação e predição de efeitos futuros da ação, este efeito é conhecido como *self-other*, e pode ser observado em outras tarefas que envolvem ação-percepção, tais como o reconhecimento de caligrafia (KNOBLICH et al., 2002) e em lançamento de dardos (KNOBLICH; FLACH, 2001). No presente estudo, foram analisados os padrões de sincronização de clarinetistas profissionais, durante a performance em um ambiente simulado de orquestra.

1.1 Gesto e música

Pesquisadores como Eduard Sievers (1850-1932), Gustav Becking (1894-1945) e Alexander Truslit (1889-1971) foram pioneiros na análise combinada de música e gesto. As demonstrações mais antigas do interesse no tema partem da proposição de que a informação gestual seria codificada na microestrutura expressiva da performance em diversos níveis estruturais (REPP, 1993). As primeiras tentativas de demonstrar empiricamente as relações entre o movimento e a música datam do início do século XX, realizadas por Gustav Becking em 1928, e por Alexander Truslit em 1938. Estes autores focaram seus trabalhos na hipótese de que a música poderia ser descrita por informações gestuais. Como apontado por Repp (1993), es-

tes trabalhos apresentam resultados de uma série de experimentos voltados para a reconstrução da informação gestual contida na música. O método se resumia na análise de curvas (posteriormente denominadas “Becking curves”) produzidas por um pequeno bastão, que era movido por um ouvinte em acompanhamento a música que era executada (NETTHEIM; BECKING, 1996).

Clynes (1995), propôs uma revisão dos experimentos de Becking fazendo o uso de um dispositivo inventado por ele próprio chamado *Sentograph*, usado para detectar variações na pressão exercida pela ponta de um dedo do ouvinte em acompanhamento a música. Assim como Becking, Clynes focou-se na descrição das nuances detectadas em performances das obras de alguns compositores, propondo o que ele chamou de ‘pulso pessoal do compositor’ (*composers’ inner pulse*) como forma de determinar curvas específicas para cada compositor estudado. Clynes argumentava que o significado da música provinha de formas ‘essenciais’ (*essentic forms*) – curvas dinâmicas que caracterizam emoções básicas, definidas pela estrutura musical, e que o ‘pulso pessoal do compositor’ deveria de alguma forma manifestar-se na microestrutura expressiva da performance (CLYNES, 1995).

No início da década de 1990, Neil Todd propôs diversos modelos de estruturação da performance, baseando-se em observações de variações de ritmos (TODD, 1989a; TODD, 1989b) e de intensidades (TODD; NEIL, 1992). Diferente de Becking e Clynes, Neil Todd concentrou-se nas emoções expressas pelo corpo humano como um todo. Nestes estudos, os dados extraídos desses modelos foram comparados com curvas de tempo e dinâmica produzidas por movimentos da cabeça do intérprete em performances de peças para piano. Os resultados indicaram uma similaridade parcial entre as curvas reais e o modelo, mas nenhuma avaliação quantitativa foi realizada (REPP, 1993; WIDMER; GOEBL, 2004). Os resultados obtidos levaram o autor a sugerir a hipótese de que existiria uma interação direta entre sistema motor e sistema auditivo, de forma que representações internas seriam evocadas diretamente no centro motor cerebral por estímulos sensoriais que correspondem especificamente a estas representações. Isto indicaria uma estreita conexão entre a expressão corporal e a expressão musical, visto que os intérpretes ouvem suas próprias performances (TODD, 1999).

Dahl e Friberg (2007) apontam que o relacionamento entre música e movimento pode ser descrito por diferentes aspectos. Destes, o mais notório é o fato de que os sons dos instrumentos acústicos tradicionais são produzidos a partir do

movimento humano, de forma que determinadas características do movimento serão inevitavelmente refletidas no som resultante.

1.2 Definição de gestos acompanhantes

Uma outra categoria de gestos pôde ser definida a partir da observação de um conjunto de movimentos não diretamente relacionados à produção do som. Estes poderiam, eventualmente, servir como via de comunicação das intenções expressivas do intérprete (CADOZ; WANDERLEY et al., 2000). Com base nas observações feitas por Delalande (1988), Cadoz, Wanderley et al. (2000) definiram duas categorias de gestos, dependentes e independentes da produção sonora: “gestos efetivos” (*effective gesture*) e “gestos acompanhantes” (*ancillary* ou *accompanist gesture*), respectivamente.

1.3 Objetivos

Este estudo tem como objetivo investigar as estratégias de sincronização utilizadas em grupos musicais, partindo de duas abordagens complementares, análise cinemática e análise de parâmetros acústicos. Estudos anteriores relataram a ocorrência do efeito *self-other* em performances musicais a partir da análise das assincronias em duos de piano (KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010). Estes resultados foram corroborados por um estudo mais recente em duos de clarineta, realizado com o mesmo conjunto de dados usados no presente trabalho (LOUREIRO et al., 2012). Os estudos citados acima se restringiram às manipulações temporais das notas, a partir de medidas de assincronia entre os inícios de cada nota tocada por cada músico.

De fato, músicos, ao tocar em conjunto, buscam ativamente “seguir” as intenções interpretativas de outros músicos. Contudo, estas intenções são percebidas não apenas a partir das manipulações temporais das notas, mas também em outras dimensões, tais como intensidades, qualidade de articulação ou colorido sonoro. Considerando a possível correlação entre os movimentos corporais dos músicos e suas intenções interpretativas, este estudo propõe responder às seguintes questões: A sincronização de parâmetros acústicos se reflete na movimentação

dos músicos? (2) Esta sincronização também se reflete em outras dimensões acústicas?

1.3.1 Justificativa

Nos últimos anos, assistimos a um rápido crescimento do interesse na relação entre o movimento e a comunicação humana. Este crescimento foi impulsionado principalmente pelo desenvolvimento de novas tecnologias para captura de movimento, o que expandiu significativamente as perspectivas dos pesquisadores nesse campo. No contexto da pesquisa em música, diversos trabalhos têm demonstrado a relevância do movimento humano como meio de comunicação de intenções e emoções na performance musical (DAHL; FRIBERG, 2007). Mas apesar da grande atenção voltada para o tema, ainda sabemos pouco sobre quais processos estariam por trás de uma possível integração entre gesto e música.

2 *Materiais e Métodos*

2.1 Escolha do instrumento

O instrumento escolhido para este estudo foi a clarineta. Esta escolha foi guiada inicialmente pela proposta de analisar movimentos que não são essenciais para a produção sonora. Recentes estudos apontam para um possível acoplamento entre esse tipo de movimento, definido como *gestos acompanhantes* por Cadoz, Wanderley et al. (2000), e a comunicação das intenções expressivas dos músicos. Na clarineta, os movimentos corporais diretamente relacionados com a produção sonora, os assim chamados *gestos efetivos* (CADOZ; WANDERLEY et al., 2000), são limitados a movimentos de pequena amplitude, tais como dos dedos, mandíbula, lábios, língua e caixa torácica. Esta característica, comum à maioria dos instrumentistas de sopro, facilita e induz o músico a se engajar em uma ampla variedade de movimentos das pernas, braços, cabeça e variações posturais, cuja independência da produção sonora é aparente.

Um outro ponto relevante para nossa escolha foi a viabilidade na aquisição de dados do movimento, os quais devem ser feitos sem dificultar a execução da tarefa proposta. Um instrumento com as características da clarineta permite a realização da coleta sem grandes perturbações na técnica dos instrumentistas, o que de outra forma poderia interferir irremediavelmente nos resultados obtidos. Outros pontos que pesaram nesta escolha foram a experiência adquirida na análise de dados acústicos produzidos pelo instrumento, como demonstrado por Loureiro et al. (2009) e Mota, Campolina e Loureiro (2009), além da grande oferta de clarinetistas experientes nas orquestras locais, disponíveis para participar dos experimentos. A combinação destes fatores apontou a clarineta como instrumento ideal para a realização deste estudo.

2.2 Participantes

Cinco músicos profissionais e um estudante foram recrutados em Belo Horizonte, Minas Gerais. Todos eles possuíam experiência anterior em prática de orquestra, e todos eles se conheciam. Dois deles tiveram aulas de clarineta com o mesmo professor. Até o momento das gravações, o estudante estava tendo aulas de clarineta com um dos outros participantes.

2.3 Material musical

Foi utilizado no experimento um excerto da “Dança do Camponês e o Urso” do balé *Petrushka* de Igor Stravinsky, extraído do *Quatrième tableau* N° 100 (três primeiros compassos), onde “o camponês toca a flauta e o urso anda sobre as patas traseiras”. Neste trecho, primeira e segunda clarineta tocam solo em uníssono (*solí a 2*), apresentando uma situação de performance real, que exige a sincronização de cada nota. A seleção do material musical foi feita segundo dois critérios: (1) linhas melódicas simples, (2) *Solí a due* - a situação de uníssono garante a igualdade de condições técnicas em ambas as tomadas.

The image shows a musical score for Clarinet in Bb. It consists of two staves. The first staff is labeled 'Sostenuto' and 'ff', and the second staff is labeled '2'. The music consists of a series of notes, with some notes marked with accents (>). The first staff has measures 1-14, and the second staff has measures 15-41.

Figura 2.1: Excerto da Dança do Camponês e o Urso do balé *Petrushka* de Igor Stravinsky, extraído do *Quatrième tableau* N° 100 (três primeiros compassos).

2.4 Experimento

As gravações foram realizadas em duas sessões, separadas por intervalos de dois a três dias. Na primeira sessão, os músicos foram instruídos a tocar como pri-

meira clarineta (*primos*), isto é, seguindo suas próprias intenções interpretativas. Uma vez terminada a sessão, foi solicitado aos músicos que escolhessem uma das quatro gravações executadas por eles. Esta gravação serviu mais tarde como *primo* para a segunda sessão.

Na segunda sessão, os músicos foram instruídos a tocar como segunda clarineta (*secondos*), acompanhando todas as gravações de *primos* escolhidas na primeira sessão, incluindo as executadas por eles mesmos. A única instrução dada foi para acompanhar o *primo* da melhor forma possível. Três batidas de metrônomo foram incluídas no início de cada gravação *primo* utilizada na segunda sessão. O andamento de cada gravação foi estimado pela extrapolação da duração dos primeiros compassos. Depois de ouvir a execução do *primo* uma vez, os clarinetistas tocaram como *secondo* quatro vezes, enquanto ouviam o *primo* através de um fone de ouvido em sua orelha direita. O uso do fone de ouvido foi necessário para impedir que o áudio do primeiro clarinetista fosse gravado junto o áudio do segundo clarinetista. Todas as gravações *primo* usadas na segunda sessão foram distribuídas aleatoriamente, de forma a evitar efeitos de ordem.

2.5 Aquisição e processamento dos dados gestuais

O sistema *Optotrak Certus*¹ foi usado para rastrear a posição tridimensional de dois corpos rígidos², a clarineta e a cabeça dos instrumentistas. Para isso foram utilizados um grupo de três marcadores em cada corpo rígido. A taxa de amostragem utilizada foi de 100Hz. Um grupo de marcadores auxiliar foi colocado em uma posição fixa em relação as câmeras, para servir como referencial de posicionamento. O primeiro grupo de marcadores foi fixado nas campanas das clarinetas e o segundo nas cabeças do participantes, de forma a permitir o rastreamento dos

¹Optotrak Certus (fabricado pela Norte Digital Inc. [NDI]) é um sistema de captura de movimento tridimensional, que utiliza diodos emissores de luz infravermelha (IREDs) como marcadores ativos. Estes marcadores são individualmente identificados por uma frequência de pulsos óticos e as suas posições são medidas por um sistema de câmeras trinocular. O sistema pode medir simultaneamente até 512 marcadores individuais ou um vasto número de corpos rígidos com marcadores embutidos.

²Corpo rígido é um conceito de representação de um corpo sólido de tamanho finito em que a deformação é desprezível. Isto é, a distância entre quaisquer dois pontos de um corpo rígido permanece constante com o tempo independentemente das forças externas exercidas sobre ele. A posição espacial de um corpo rígido pode ser representada pelas coordenadas de um ponto de referência do corpo (normalmente coincidente com a centróide do sólido), e de sua posição angular, descrita pela rotação em torno dos eixos de referência.

seis graus de liberdade dos corpos com o mínimo de interferência na movimentação do músico – para mais detalhes e exemplos de configurações em experimentos similares, ver Wanderley et al. (2005), Keller, Knoblich e Repp (2007), Keller e Appel (2010).

Estudos anteriores demonstraram que a movimentação da campana da clarineta é responsável por grande parte do movimento realizado durante a performance do instrumentista (WANDERLEY, 2002; PALMER et al., 2009). Em Wanderley et al. (2005) os movimentos do corpo dos clarinetistas e da clarineta foram quantificados a partir do cálculo das diferenças entre quadros sucessivos de gravações de vídeo. Os resultados demonstraram que 30% da movimentação total foi realizada pela campana, contra 20% para os movimentos da cabeça e 10% para os movimentos das pernas. Com base nestes resultados optamos por restringir este estudo à análise do movimento da campana.

2.5.1 Parametrização

Foi definido um parâmetro escalar a partir das posições tridimensionais, denominado *perfil de velocidade* das execuções. Este parâmetro foi calculado como a velocidade instantânea da centróide dos marcadores acoplados à campana da clarineta:

$$v_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2} \quad (2.1)$$

onde x , y e z representam as coordenadas espaciais, e i o número da amostra.

2.5.2 Definição da região de análise

Tendo como objetivo testar a hipótese de acoplamento entre parâmetros acústicos e a movimentação dos músicos, procuramos restringir a análise ao domínio temporal correspondente ao evento musical, tomando cuidado para não descartar informação eventualmente importante logo antes e logo depois da execução musical, tal como gestos de preparação e/ou finalização. Adotamos como limitador inicial o instante correspondente a um pulso métrico antes do onset da primeira nota e como limitador final o instante correspondente a um pulso métrico após o

offset da última nota. A estimação do andamento da execução foi feita a partir do pulso calculado pelas notas 5 e 9 (Fá# e Dó), já que o excerto começa na segunda parte do primeiro tempo, o que dificulta estabelecimento imediato do andamento, mesmo para os músicos mais experientes.

2.5.3 Suavização dos *perfis de velocidade*

O sistema de captura de movimento utilizado na aquisição dos dados fornece uma precisão de até 0.1mm que, aliada à taxa de amostragem de 100Hz utilizada no experimento, resulta em uma alta resolução da coleta dos dados. Com isto são capturados os movimentos mais sutis realizados pelos músicos, tais como aqueles causados pelo impacto dos dedos no instrumento. Para os propósitos deste experimento este tipo de movimento é considerado como ruído, pois nenhuma informação pode ser extraída deles. Portanto eles foram minimizados com um filtro Butterworth passa-baixas, com ordem 3 e frequência de corte em 5Hz.

2.5.4 Normalização da amplitude

Os participantes do experimento possuíam diferentes alturas, composições corporais e idades. Isto nos fornece resultados com grande variabilidade, pois alguns participantes tendem a realizar movimentos mais amplos e mais velozes que os outros. Para evitar grandes diferenças e fornecer o material adequado para a comparação das performances, realizamos a normalização destas amplitudes. O *perfil de velocidade* de cada execução teve sua amplitude normalizada, de tal maneira que seu valor RMS (*Root Mean Square*) seja igual a 1. O RMS é definido como:

$$v_{RMS} = \sqrt{\frac{1}{n}(v_1^2 + v_i^2 + \dots + v_n^2)} \quad (2.2)$$

onde v representa o *perfil de velocidade* e n o número da amostra. Esta normalização é realizada a fim de garantir uma comparação mais adequada dos perfis de músicos de massa corporal diferentes.

2.5.5 Ajustamento Temporal

A manipulação de parâmetros temporais da música foi amplamente empregada pelos sujeitos durante as gravações. Como visto anteriormente, isso significa uma grande variabilidade nos perfis temporais das gravações, mesmo quando executadas pelos mesmos intérpretes. A Figura 2.2 apresenta a superposição de dois *perfis de velocidade*, executados pelo mesmo intérprete em sessões de gravação distintas. Podemos constatar que a discrepância entre os dois sinais se intensifica ao longo do tempo. Isto é devido a uma pequena variação de andamento existente entre as duas execuções, que resulta em uma diferença de duração de aproximadamente 400ms entre elas.

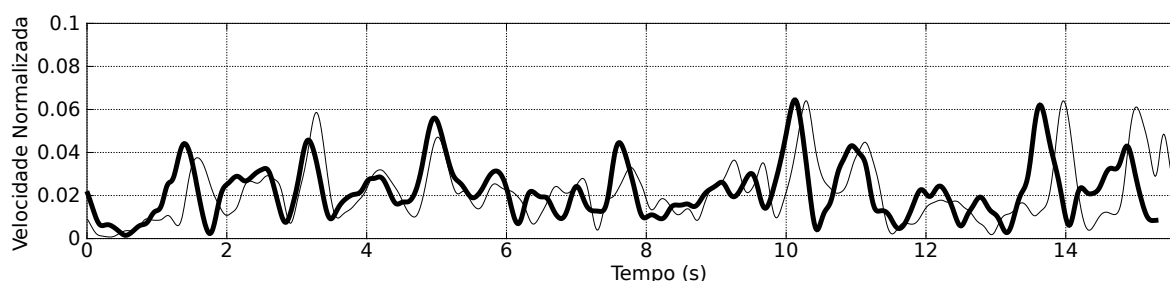


Figura 2.2: Superposição de dois *perfis de velocidade* executados com diferentes andamentos. A curva mais clara possui cerca de 400ms a mais de duração em relação à curva mais escura.

Como mostra a Figura 2.2, a diferença entre as durações é responsável por um aumento gradativo da defasagem entre os picos de velocidade. Os instantes onde estes picos ocorrem variam de acordo com os andamentos adotados pelo intérprete, o que dificulta a comparação entre execuções com diferentes andamentos ou variações temporais (*rubato*, *accelerando*, etc.). Visando minimizar o desalinhamento destes picos, utilizamos a técnica de *time warping*, como sugerido por Wanderley (2002) e Wanderley et al. (2005) para o ajustamento temporal dos sinais, de forma a garantir o mesmo andamento aproximado para todas as execuções. Considerando a interdependência entre o movimento do músico e o que ele toca, um ajuste temporal dos sinais de forma a otimizar o alinhamento das notas, possibilita uma melhor correspondência dos movimentos com a estrutura musical, permitindo assim uma análise comparativa mais adequada entre os movimentos das diferentes execuções.

O primeiro passo no processo de ajustamento temporal é a definição de um

modelo temporal, que vai servir de base para a normalização de todas as execuções. Uma opção é a criação de um modelo sintético, construído a partir dos valores extraídos da partitura. Outra possibilidade é usar a média dos valores dos *onsets* de todas as gravações para criar o modelo. A segunda opção possui como vantagem a representação das intenções temporais dos instrumentistas. Neste estudo, optamos pela segunda possibilidade.

O modelo temporal utilizado no processo de *time-warping* é definido por uma sequência de instantes temporais (início de cada nota) calculados como a média dos inícios de uma dada nota, extraída de todas as execuções. Portanto, este resultado pode ser visto como o *perfil temporal médio* deste conjunto de dados. Os *perfis de velocidade* são então reamostrados entre os valores de cada instante temporal fornecido por este modelo, utilizando interpolação por splines cúbicos.

A Figura 2.3 mostra o resultado do *time warping* aplicado às mesmas curvas apresentadas na Figura 2.2. Podemos verificar que o ajuste temporal melhora o alinhamento dos picos dos *perfis de velocidade*. Isto sugere que a movimentação do clarinetista segue algum tipo de organização intrínseca que faz com que o músico realize determinados movimentos em posições específicas da partitura, uma vez que os *perfis de velocidade* foram ajustados em relação às notas executadas. Uma hipótese levantada, é a de que esta organização poderia ser ditada por algum elemento estrutural da música (andamentos, ritmos, perfil melódico, perfil de intensidades, etc.).

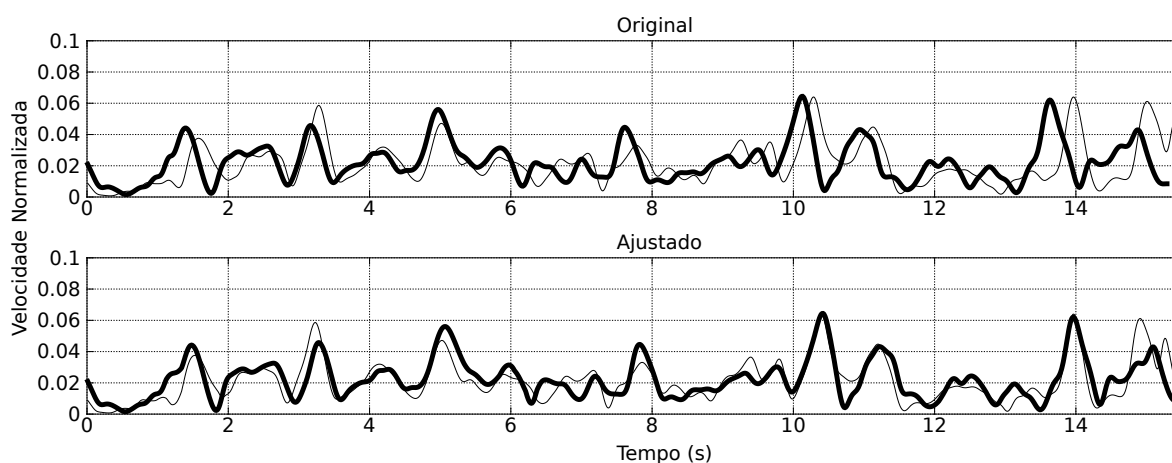


Figura 2.3: Ajustamento temporal de *perfis de velocidade* executados com diferentes andamentos. O painel superior apresenta as curvas originais, enquanto o painel inferior apresenta as curvas ajustadas.

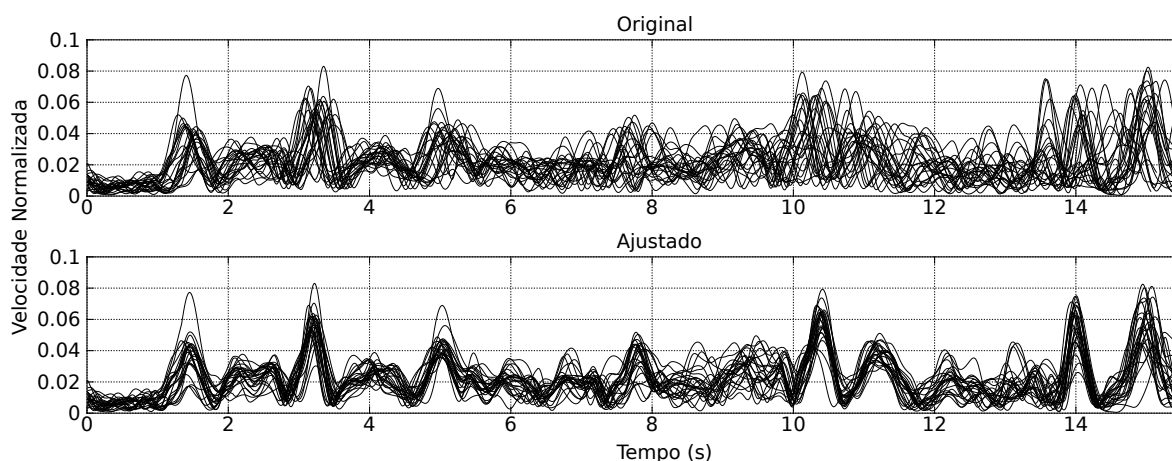


Figura 2.4: Superposição de 24 *perfis de velocidade* executados pelo mesmo músico, antes e depois do ajustamento temporal.

Aplicamos esse procedimento em todas as gravações do nosso conjunto de dados. Isto permitiu a comparação de gravações feitas sob diferentes restrições temporais, que no caso deste estudo, são ditadas pelo clarinetista *primo*. O alinhamento temporal dos *perfis de velocidade*, após a aplicação do *time warping* em 24 execuções feitas por um mesmo intérprete (4 vezes acompanhando cada um dos 6 primeiros), pode ser visto na Figura 2.4.

2.6 Aquisição e processamento dos dados acústicos

2.6.1 Gravação

O áudio foi capturado utilizando uma interface M-Audio FireWire 1814, gravado com frequência de amostragem 44.100 Hz, em apenas um canal, utilizando um microfone omnidirecional numa sala com tratamento acústico básico. O microfone foi colocado a um metro de distância dos sujeitos, a fim de evitar mudanças na amplitude do sinal de som, causadas pelo movimento dos clarinetistas. Cada clarinetista utilizou o seu próprio instrumento durante as sessões de gravação. O áudio captado foi simultaneamente direcionado para a unidade de aquisição de dados do sistema Optotrak (ODAU – *Optotrak Data Acquisition Unit II*). Trata-se de um conversor analógico digital utilizado para a captura de sinais externos como os provenientes de sensores de EMG, placas de força, acelerômetros e medidores de tensão. Este sistema garante que os dados externos capturados se-

jam sincronizados com os dados de posição captados pelas câmeras do Optotrak. Para isso, dois canais foram utilizados: um com o áudio do primeiro clarinetista e o outro com o áudio do segundo clarinetista. A captação do áudio pelo ODAU foi necessária para garantir a sincronia dos dados de áudio com os captados pelo Optotrak.

2.6.2 Segmentação

O passo seguinte foi a segmentação das faixas de áudio, que foi realizada automaticamente no nível de notas, a partir da detecção dos instantes de início da nota (*onset*), final de ataque (*attack*), início de decaimento (*release*) e final da nota (*offset*). Portanto, a configuração espectral dos segmentos das notas carregam informação significativa sobre os parâmetros musicais, tais como a qualidade das transições entre notas consecutivas, articulações e agrupamentos de notas.

A segmentação de sinais musicais tem sido o foco de um grande número de estudos nos últimos anos. A revisão realizada por Bello et al. (2005) demonstra que a detecção de variações abruptas no envelope de amplitude é o método mais comum utilizado para esse fim. Essa abordagem é muito adequada para sons percussivos, como os de um piano ou um instrumento de cordas dedilhadas como o violão, mas muito ineficiente para instrumentos com envelope de energia controlável, como instrumentos de sopro (clarineta, oboé, trompete) ou de cordas friccionadas (violino, violoncelo, contrabaixo). Para solucionar este problema, consideramos o uso de um sistema de detecção de instantes temporais híbrido, desenvolvido em estudo anterior (LOUREIRO et al., 2009). Este sistema utiliza uma combinação de parâmetros espectrais e temporais do sinal de áudio: a detecção de variações no RMS janelado do sinal, e a detecção de alterações superiores a 6% nos valores de altura (*pitch*).

2.6.3 Parametrização do áudio

A parametrização do áudio se baseia na extração de informação musical do sinal de áudio, pertinente à estrutura musical (estrutura rítmica, tessitura, etc) e da performance (articulações, variações de timbres e andamentos, etc). Estas características, comumente denominadas *descritores acústicos*, são extraídas do sinal

em níveis de segmentação apropriados aos objetivos da análise, que no caso deste estudo foi o nível das notas. Ou seja, cada execução fornece 41 valores para cada descritor, o número de notas do excerto.

Descritores acústicos

A busca por um conjunto de descritores acústicos que sejam adequados para o mapeamento das intenções interpretativas dos músicos é ainda um problema em aberto. Até o momento atual, foram realizadas diversas tentativas para criar um modelo capaz de descrever a forma como os músicos comunicam suas intenções expressivas. A maior parte destes estudos tem como foco o desenvolvimento de sistemas de renderização de performances expressivas, ou sistemas de indexação musical para mecanismos de busca (MION; POLI, 2008). Por exemplo Maestre et al. (2006) utiliza uma combinação de descritores para gerar sequências musicais expressivas a partir de um arquivo MIDI, Molina-Solana, Arcos e Gomez (2008) propõe a identificação automática de violinistas profissionais usando o contorno melódico e a variação de energia das execuções. Uma revisão ampla que inclui a análise de diversos modelos pode ser encontrado em Widmer e Goebel (2004). Uma revisão dos descritores acústicos mais utilizados em tarefas de classificação musical pode ser encontrada em Peeters (2004) e Peeters e Rodet (2002).

Um dado conjunto de descritores pode ter uma performance excepcional para um determinada tarefa, mas resultar em resultados insatisfatórios em outra. Diferentes propósitos pedem diferentes conjuntos de descritores. A seleção dos descritores mais adequados pode ser um problema, visto que o número de descritores propostos pelos sistemas de classificação modernos pode facilmente atingir uma centena. Neste estudo a escolha dos descritores foi guiada pela busca de uma representação, ainda que rudimentar, dos parâmetros musicais mais relevantes para o excerto. Escolhemos portanto um descritor para representar cada parâmetro musical. Detalhamos a seguir os descritores acústicos utilizados na análise.

IOI normalizado (perfil temporal)

Como demonstrado por Repp (1995), o perfil temporal de uma performance musical é uma das principais fontes de variações expressivas utilizadas pelos intérpretes. O valor nominal das durações das notas, registrados na partitura, são

utilizados para normalizar os valores de IOIs (*Inter Onset Interval*)³ das diferentes execuções gravadas. Este procedimento permite a criação de um “perfil temporal” de cada performance.

Índice de articulação

Alguns aspectos da qualidade das transições entre notas estão relacionadas à manipulação intencional de articulação pelo intérprete, através do controle preciso das durações e da qualidade dos ataques, que ele detém. A articulação está, portanto, estreitamente relacionada com as intenções expressivas e de inteligibilidade do intérprete. Conforme as regras de articulação definidas em Bresin (2001), implementamos um índice de articulação, definido como a razão entre a duração da nota DR (intervalo de tempo entre o início e o final da nota) e o intervalo de tempo entre seu início e o início da nota subsequente:

$$AI(i) = 1 - \frac{DR(i)}{IOI(i)} \quad (2.3)$$

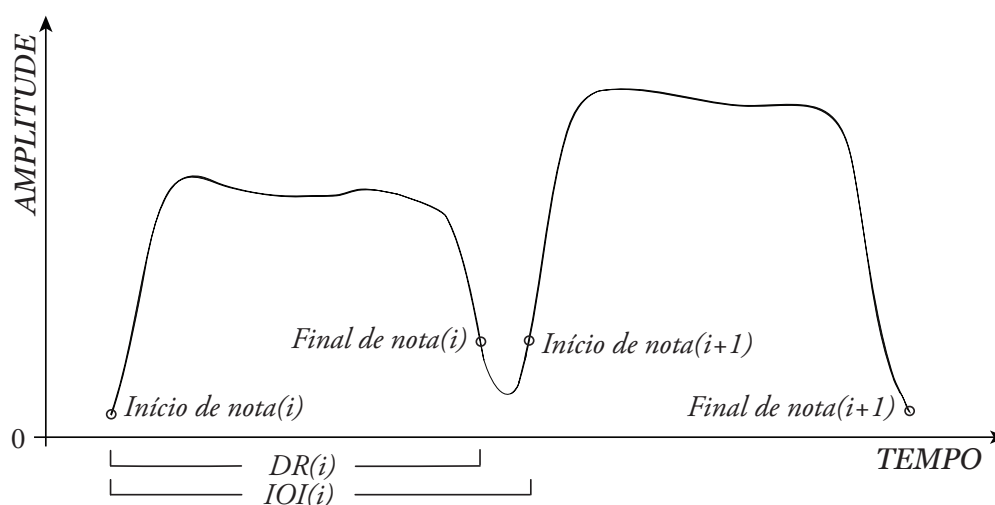


Figura 2.5: Representação do cálculo do Índice de Articulação para a transição entre duas notas.

Este índice é apropriado para descrever as transições entre notas destacadas, normalmente produzidas na clarineta com interrupções bruscas do fluxo de ar

³Este termo se refere à quantidade de tempo entre o início de uma nota e o início da nota seguinte. (PARNCUTT, 1994).

através de golpes da língua na palheta, com os quais o intérprete controla a qualidade dos ataques e a duração de cada nota. Vale ressaltar que a habilidade do intérprete em controlar a duração das notas depende muito das condições de reverberação do ambiente. Tendo seu valor próximo a 0 para transições em *legato*, este índice não se apresentou adequado para descrever estas transições.

Índice de Legato

Na clarineta as notas com a articulação *legato* são produzidas com um sopro único sem interrupção do fluxo de ar, ao longo de todo o trecho com ligadura. Para investigar a qualidade das transições entre as notas executadas com esta intenção, definimos o *Índice Legato*, como sugerido por Maestre e Gómez (2005), a partir de um *legato* ideal entre duas notas sem qualquer decréscimo de energia, representado pela reta L , definida pelos instantes de início de decaimento de uma nota e final do ataque da nota subsequente. O índice é estimado pela razão entre a área $A1$, compreendida entre esta reta e a curva de energia na região de transição entre as notas e a área total abaixo desta reta, $A1 + A2$:

$$IL(i) = 1 - \left[\frac{A_1}{A_1 + A_2} \right] = 1 - \left[\frac{\sum_{t=id(i)}^{fa(i+1)} L(t) - RMS(t)}{\sum_{t=id}^{fa(i+1)} L(t)} \right] \quad (2.4)$$

onde id é o início do decaimento e fa o final do ataque.

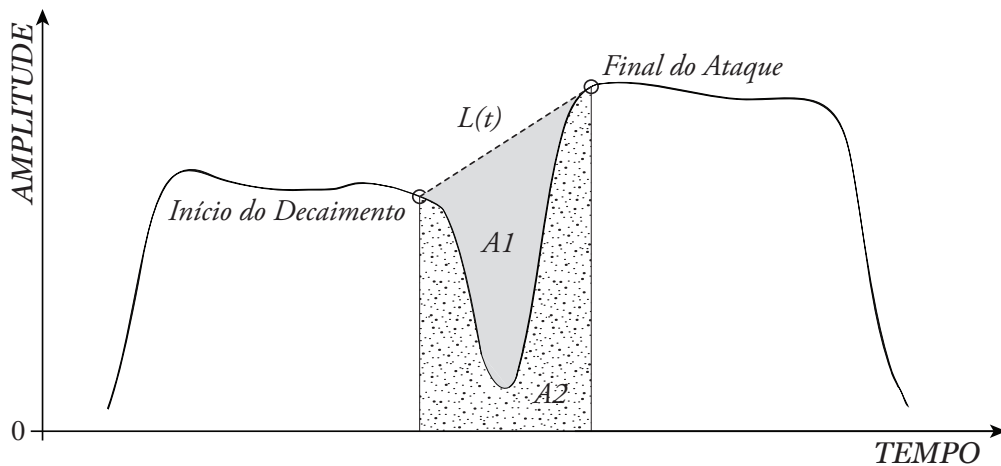


Figura 2.6: Cálculo do *Índice de Legato* para a transição entre duas duas notas.

Logaritmo do tempo de ataque

Devido à sua multidimensionalidade, o timbre é um dos parâmetros mais complexos para se descrever, por não existir ainda uma maneira objetiva para sua normalização. Para descrever o timbre, adotamos o *Logaritmo do Tempo de Ataque*, um parâmetro, cuja relação com a percepção de timbre foi proposta por McAdams (1999) e que é manipulada pelo intérprete a partir da qualidade do ataque de cada nota.

Ataque é definido como o período de tempo necessário para um som atingir a sua amplitude inicial máxima. Ele tende a ser maior em instrumentos de excitação contínua e muito curto para sons percussivos (KIM; MOREAU; SIKORA, 2005). O ataque pode ser caracterizado não apenas pela sua duração, mas também pela sua inclinação. A região de ataque é caracterizada por variações rápidas no espectro de frequências, e na clarineta ele influencia principalmente a percepção dos timbres. O *Logaritmo do tempo de ataque* é definido como o logaritmo (em base decimal) da duração entre o início da nota T_i e o final do ataque T_f :

$$LTA = \log_{10}(T_f - T_i) \quad (2.5)$$

Loudness

Loudness é um atributo psicoacústico relacionado com a sensação de intensidade percebida de um som. Como um parâmetro musical este atributo é referenciado como *dinâmica*. O conceito de *loudness* foi criado para caracterizar a sensação de volume de qualquer som. Foi introduzido na década de vinte por Heinrich Barkhausen (FASTL; ZWICKER, 2007). A relação de intensidade entre dois sons é normalmente medida em decibéis, o qual é definido como a relação logarítmica entre 2 níveis, um dos quais é um nível de referência. (MOORE, 2003, p. 8-9). Os valores de *loudness* foram calculados pela média do RMS do sinal de de cada nota, para isso utilizamos a convenção de que um tom puro a 100dB possui um valor médio RMS igual a 1, seguindo o modelo definido por Jepsen, Ewert e Dau (2008, p. 425).

A percepção do *loudness* também é influenciada pela altura da nota (FASTL; ZWICKER, 2007). Por isso, após a transformação para *dB SLP* é necessário ponderar os valores de *loudness* considerando a resposta auditiva humana. Esta ponde-

ração é realizada utilizando um filtro (A-weighted), especificado pela norma ANSI/ASA S1.42-2001 regida pelo Instituto Americano de Normas Nacionais (ANSI) (RAICHEL, 2006). Este filtro de ponderação baseia-se na curva de intensidades iguais (*equal-loudness*) definida nos estudos realizados por Fletcher-Munson (FASTL; ZWICKER, 2007).

2.6.4 Exclusão de valores inválidos

A análise proposta se baseia na comparação entre os valores de descritores relativos a cada nota executada pelos intérpretes. A viabilidade deste estudo está ligada à correta associação das notas entre diferentes performances. Durante as sessões de gravação erros na execução de algumas notas foram cometidos pelos intérpretes (notas que simplesmente não foram tocadas ou com resultado ruidoso, conhecidos como “guincho”), inviabilizando desta forma a utilização destas no estudo. Uma nota descartada em uma execução implica na exclusão de todas as notas correspondentes em todas as outras execuções. No total foram excluídas 17% de todas as notas.

2.7 Análise dos dados

Como descrito anteriormente, as gravações utilizadas neste estudo foram realizadas simulando uma situação de orquestra, onde o segundo clarinetista deve seguir as intenções interpretativas do primeiro clarinetista. Partindo deste princípio, definimos dois agrupamentos: *secondos* e *primos* que foram utilizados para classificar as diferentes execuções.

O agrupamento por *secondos* diz respeito a todas as execuções realizadas por um mesmo músico, independente de qual *primo* ele estava acompanhando. Por exemplo, o músico **C1** realizou 4 gravações acompanhando cada um dos outros 5 músicos, e a si próprio, resultando em 24 gravações do músico **C1** como segundo clarinetista.

O agrupamento por *primos* relaciona as gravações da primeira sessão do experimento aos primeiros clarinetistas que foram acompanhados. O objetivo desta organização foi encontrar características nos *perfis de velocidade* nas gravações da

segunda sessão do experimento, que pudessem ser comuns a todas as gravações dos segundos clarinetistas, quando eles acompanharam um determinado *primo* da primeira sessão.

2.7.1 Análise Discriminante Linear

A Análise Discriminante Linear (*Linear Discriminant Analysis* - LDA) foi utilizada para a separação destes agrupamentos propostos, considerando cada amostra como uma dimensão do espaço dos dados. A técnica propõe a transformação linear de observações multivariadas \mathbf{x} em observações univariadas \mathbf{y} , pertencentes a uma população π_i , $i = 1, 2, 3, \dots, k$, de tal forma que sejam o mais separadas o possível dos valores de \mathbf{y} pertencentes à população π_j , para $j \neq i$ (JOHNSON; WICHERN, 1998, p. 629). O conceito de discriminação eventualmente se superpõe ao conceito de classificação, que é definido como a alocação de objetos em classes previamente identificadas.

A técnica LDA é aplicada primeiramente para treinar um classificador com os dados observados dos diferentes grupos, para em seguida classificar novas observações (JOHNSON; WICHERN, 1998, p. 683). Neste estudo utilizamos a LDA para classificar os dados acústicos e cinemáticos extraídos de cada execução em dois diferentes agrupamentos, clarinetistas *primo* e clarinetistas *secondo*. Posteriormente a performance deste classificador foi testada através de uma validação-cruzada (DUDA; HART, 1996, cap. 9, p. 33). A LDA foi realizada utilizando o pacote MASS (VENABLES; RIPLEY, 2002) do ambiente de computação estatística R (R Development Core Team, 2008).

Considerando \mathbf{x} uma dada observação, com média μ , extraída de uma população com k grupos distintos, a LDA encontra a combinação linear de \mathbf{x} , expressa pela matriz \mathbf{U} , que proporciona a maior separação entre os grupos sugeridos (FISHER, 1938; DUDA; HART, 1996):

$$\mathbf{y} = \mathbf{U}(\mathbf{x} - \mu) \quad (2.6)$$

O cálculo da matriz \mathbf{U} depende da forma como os dados estão agrupados, portanto, $k - 1$ discriminantes lineares (LDs) vão ser calculados, aparecendo então nas colunas da matriz \mathbf{U} .

2.7.2 Redução da Dimensionalidade: Análise em Componentes Principais

Os resultados da LDA podem ser fortemente influenciados pelo número de variáveis (dimensões) amostradas para cada observação. Estes efeitos são os chamados “problemas da dimensionalidade” ou “maldição da dimensionalidade”, como apontado por Duda e Hart (1996, capítulo 3, pág. 27). Este problema pode ser frequentemente observado em situações práticas onde, além de um certo ponto, a inclusão de dimensões adicionais pode prejudicar, ao invés de ajudar, o processo de discriminação dos grupos.

Dado o número elevado de dimensões (1500 amostras), a LDA foi aplicada em combinação com a Análise em Componentes Principais (PCA) para a redução da dimensionalidade dos dados. O objetivo da Análise em Componentes Principais (PCA) é explicar a estrutura de variância-covariância de um conjunto de variáveis. Isto é feito através de uma combinação linear destas variáveis. Ou seja, a PCA encontra a combinação linear entre as variáveis que melhor explica a variância entre elas (JOHNSON; WICHERN, 1998, p. 458).

Para encontrar o número de PCs ideal para a discriminação entre os grupos propostos, a LDA é calculada recursivamente, iterando o número de PCs utilizados. Para cada iteração calculamos o número de classificações erradas usando uma técnica de MAP (máxima probabilidade posterior) nos valores previstos pelo modelo LDA. Uma representação gráfica do procedimento pode ser vista na Figura 2.7. O uso combinado das técnicas PCA e LDA permite portanto reduzir o número de dimensões utilizadas na análise e ao mesmo tempo revelar quais as variáveis são mais importantes para a discriminação de um dado grupo de observações. A combinação destas técnicas não é nova, e tem seu uso comumente aplicado em sistemas de reconhecimento de face, como por exemplo, os relatados em Marcialis e Roli (2006) e Lu, Wang e Jain (2003)

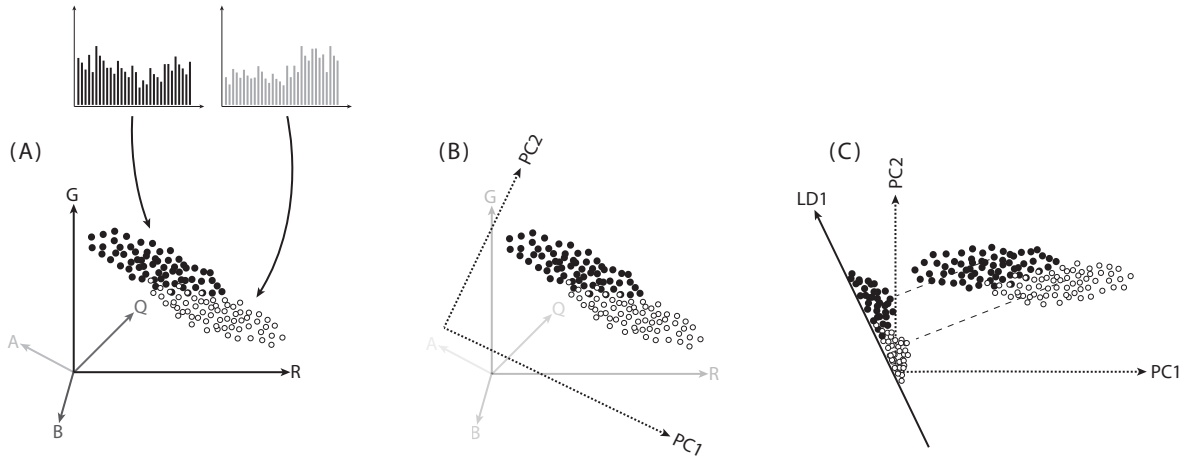


Figura 2.7: Representação ilustrativa do uso combinado PCA-LDA. (A) Cada ponto amostrado nos *perfis de velocidade* é tratado como uma dimensão num espaço. (B) A PCA projeta os dados nos vetores com maior variância explicada. (C) A LDA projeta os dados no vetor que maximiza a separação entre as classes.

O cálculo combinado PCA-LDA retorna o total de classificações erradas em relação ao número de PCs considerados. Usamos esse resultado para selecionar o número ideal de PCs necessário para uma classificação sem erros.

2.7.3 Retro-Projeção

O efeito de um dado componente LD no espaço dimensional original, pode ser obtido através da pseudo-inversa da matriz \mathbf{U} . Este procedimento encontra o vetor de mínima norma, no espaço original, que corresponde ao mesmo vetor no espaço LD. Considere um conjunto de dados dividido em seis grupos, o cálculo da retro-projeção do componente LD2 calculado para este conjunto é dado por:

$$\mathbf{x} - \boldsymbol{\mu} = \mathbf{U}^+ [0, 1, 0, 0, 0]^T \quad (2.7)$$

No caso do uso combinado PCA-LDA precisamos levar em conta o fato de que os dados foram previamente rotacionados pela PCA:

$$\mathbf{y} = \mathbf{UR}(\mathbf{x} - \boldsymbol{\mu}) \quad (2.8)$$

onde \mathbf{R} é a matriz de rotação da PCA, após a redução dimensional de \mathbf{x} . A pseu-

do inversa da transformação $[\mathbf{UR}]^+$ é portanto $\mathbf{R}^T \mathbf{U}^+$. Utilizamos esta matriz para retro-projetar os vetores LDs no espaço dimensional original. Este procedimento pode ainda ser utilizado como métrica para quantificar a contribuição de cada variável (dimensão) para a classificação realizada. Para tanto, calculamos um vetor de significância s da seguinte forma:

$$s_i = |\mathbf{R}^T \mathbf{U}^+(\bar{x}_i - \mu)| \quad (i = 1, \dots, r) \quad (2.9)$$

onde \bar{x} é média dos discriminantes para uma dada classe, μ a média da população e r o número de dimensões no espaço original dos dados.

3 *Resultados*

3.1 **Discriminação dos dados gestuais**

A existência de recorrência entre as execuções de um mesmo intérprete pôde ser demonstrada com a aplicação do *time warping*, como mostram as figuras 2.3 e 2.4 da Sessão 2.5.5. Da mesma forma, foi constatado que cada músico possui um perfil de movimentação com características próprias e diferentes dos perfis dos outros músicos. Portanto, podemos esperar que este agrupamento possua uma boa taxa de discriminância.

3.1.1 **Agrupamento por *secondos***

O resultado do LDA para o agrupamento por *secondos*, utilizando o espaço temporal integral de 1500 dimensões, é mostrado na Figura 3.1. O excelente resultado da discriminação não é surpreendente, pois cada músico possui uma forma própria de se mover, independente de qual *primo* eles estão seguindo, corroborando o que o alinhamento dos picos do *perfil de velocidade*, obtido com o ajustamento time-warping nos mostrou na Sessão 2.5.5. No entanto, a elevada dimensionalidade (1500) dos dados, comparada ao conjunto reduzido de 130 observações, não permite uma análise apropriada da discriminação desejada - a LDA busca uma combinação linear que melhor represente a separação entre os grupos sugeridos, a partir de diferenças não necessariamente significativas.

Isto pode ser verificado na Figura 3.2, onde são exibidas as alterações nos *perfis de velocidade* projetadas no componente LD2 o qual, segundo o resultado mostrado na Figura 3.1, é o principal responsável pela discriminação entre o clarinetista C1 e os outros. A “maldição da dimensionalidade” pode ser diagnosticada pelo comportamento altamente oscilatório desse vetor.

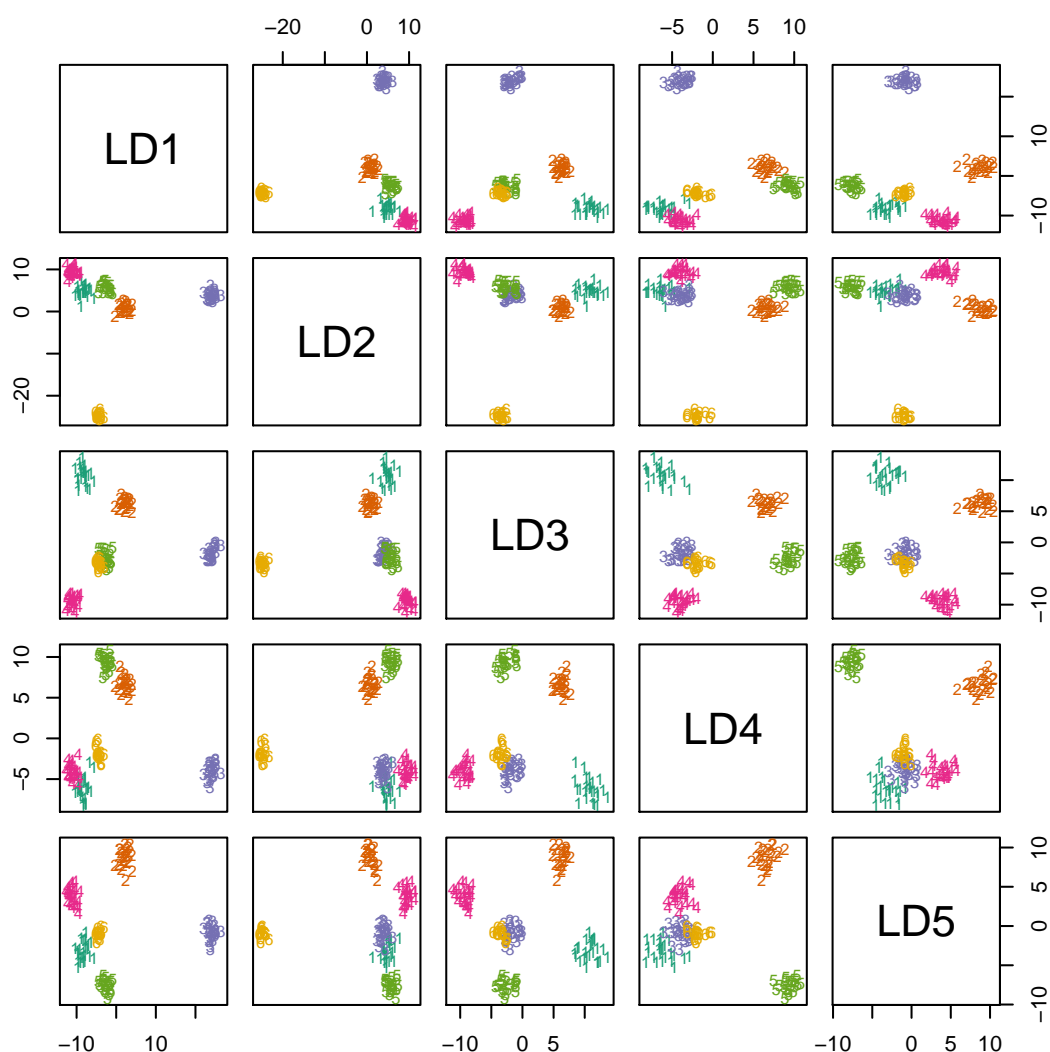


Figura 3.1: Resultados do LDA para o agrupamento por *seconds*. Os valores calculados para o espaço de LDs são mostrados em espaços de duas dimensões para todas as combinações de componentes LD. As execuções são indicadas com o rótulo de cada músico.

A redução do espaço dimensional foi realizada utilizando a combinação PCA-LDA, descrita na Sessão 2.7.3. O resultado da aplicação deste procedimento ao agrupamento por *seconds* é mostrado na Figura 3.3, onde é representado o número classificações erradas conjuntamente com a quantidade de variância acumulada dos PCs. A Figura 3.3 mostra que para este agrupamento apenas 10 PCs foram necessários para uma separação completa dos grupos.

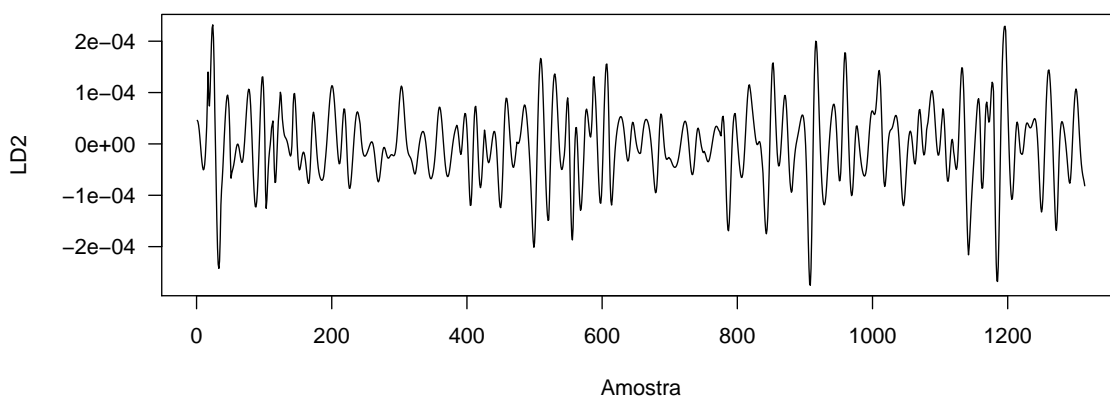


Figura 3.2: Retro-projeção do componente LD2 calculado com o espaço dimensional integral. O comportamento ruidoso deste componente indica que variações sutis entre os grupos estão sendo responsáveis pela discriminação.

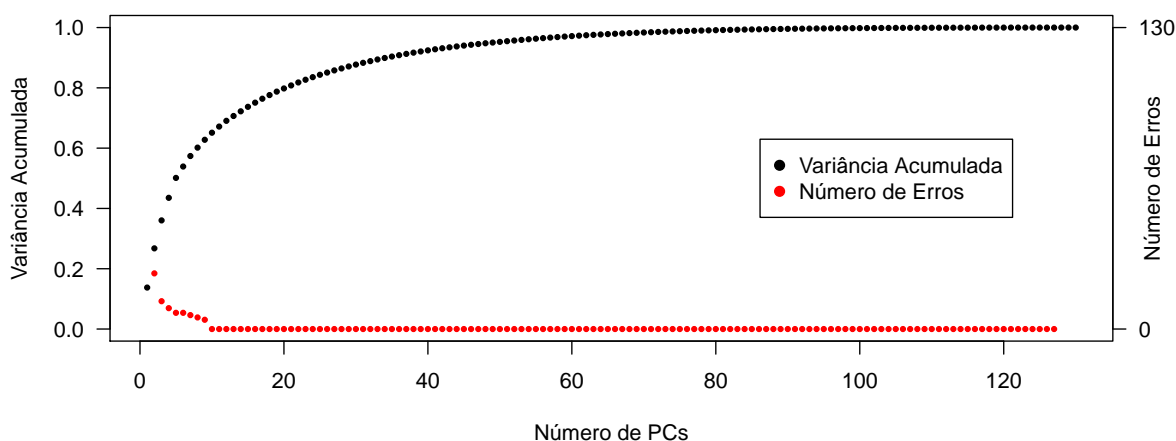


Figura 3.3: Número de classificações erradas para a análise combinada PCA-LDA para o agrupamento por *secondo*, mostrado em vermelho. A quantidade de variância acumulada é mostrada em preto.

O resultado da combinação ótima de PCs para a discriminação do agrupamento por *secondos* é mostrado na Figura 3.4. Como era esperado, os grupos classificados pela combinação das duas técnicas acabam menos separados em comparação com a discriminação resultante da LDA calculado com o espaço dimensional integral (ver Figura 3.1).

Contribuição dos componentes LD para cada *secondo*

Cada clarinetista, tocando como segundo, vai ocupar uma determinada região no espaço de componentes LDs, como visto na Figura 3.4. A contribuição de cada

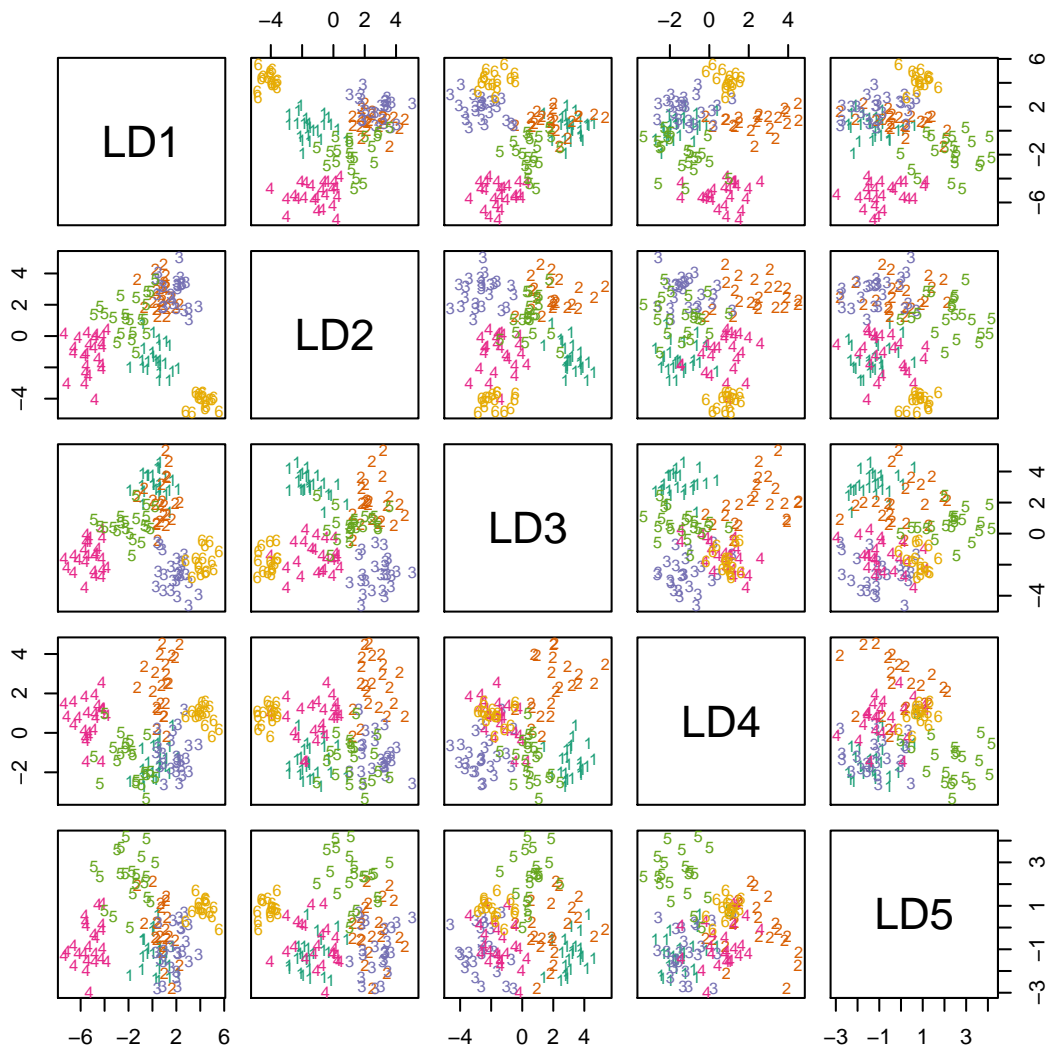


Figura 3.4: Resultados da LDA calculada para o agrupamento por *segundos* a partir dos dados transformados pela PCA.

componente LD para cada *segundo* pode ser visualizada na Figura 3.5. Os valores previstos no espaço LD para as execuções foram ponderados de acordo com cada grupo de *segundos*. As médias dos valores são representadas por pontos de tamanhos e cores variados. Valores positivos são representados em verde e valores negativos em vermelho. O tamanho e a transparência de cada ponto é proporcional ao valor médio ao longo de cada componente LD, normalizado pelo valor mais extremo para o LD respectivo na população (o máximo no caso dos pontos verdes ou o mínimo no caso dos pontos vermelhos).

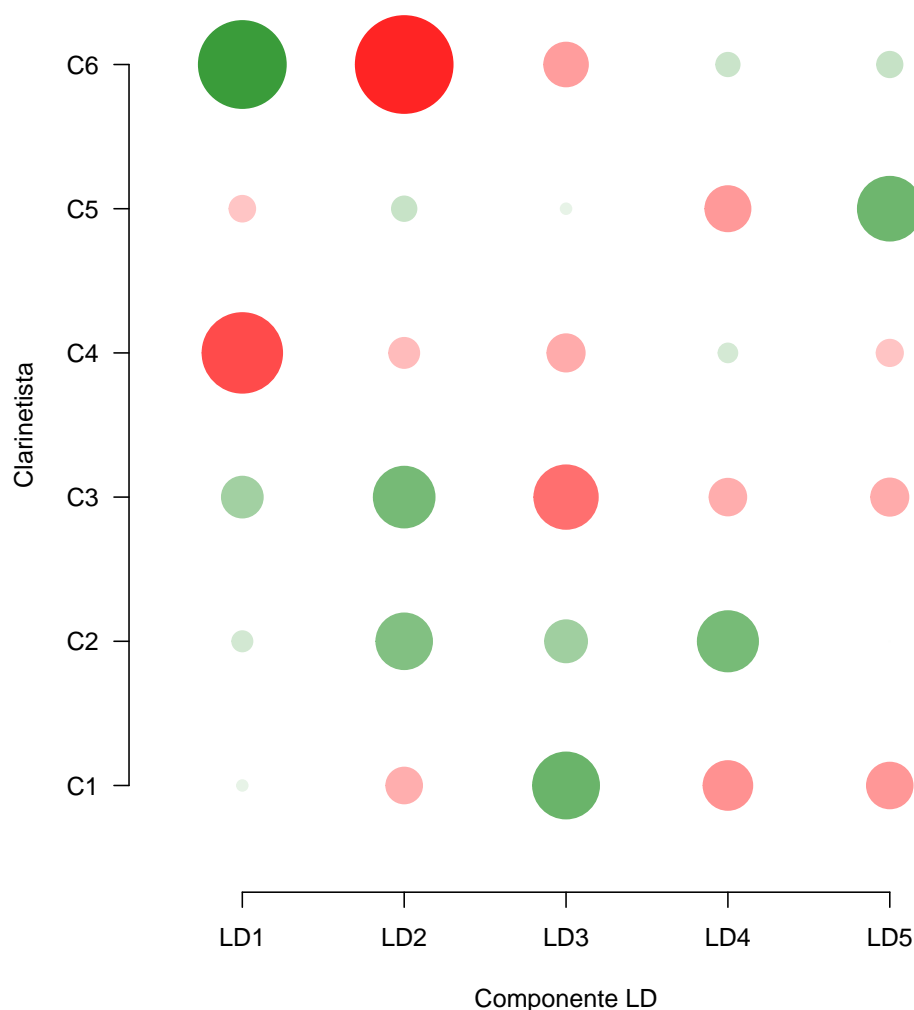


Figura 3.5: Contribuição de cada componente LD para cada clarinetista *secondo*.

Retro-projeção dos componentes LD - Agrupamento por *Secondos*

Como descrito na Sessão 2.7.3, a retro-projeção dos LDs no espaço dimensional original permite a identificação dos períodos temporais mais relevantes para a discriminação entre os grupos. Para exemplificar esse efeito, vamos analisar individualmente os resultados da retro-projeção dos componentes LD para o clarinetista **C4**. O painel superior da Figura 3.6 mostra a superposição dos *perfis de velocidade* de todas as execuções gravadas pelo clarinetista **C4**, conjuntamente com a média desses valores. O painel central apresenta todas as execuções de todos os clarinetistas superpostas, juntamente com sua média. Finalmente, o painel inferior destaca a retro-projeção do vetor com a média dos valores dos componentes

LDs para o clarinetista **C4** no espaço original dos dados.

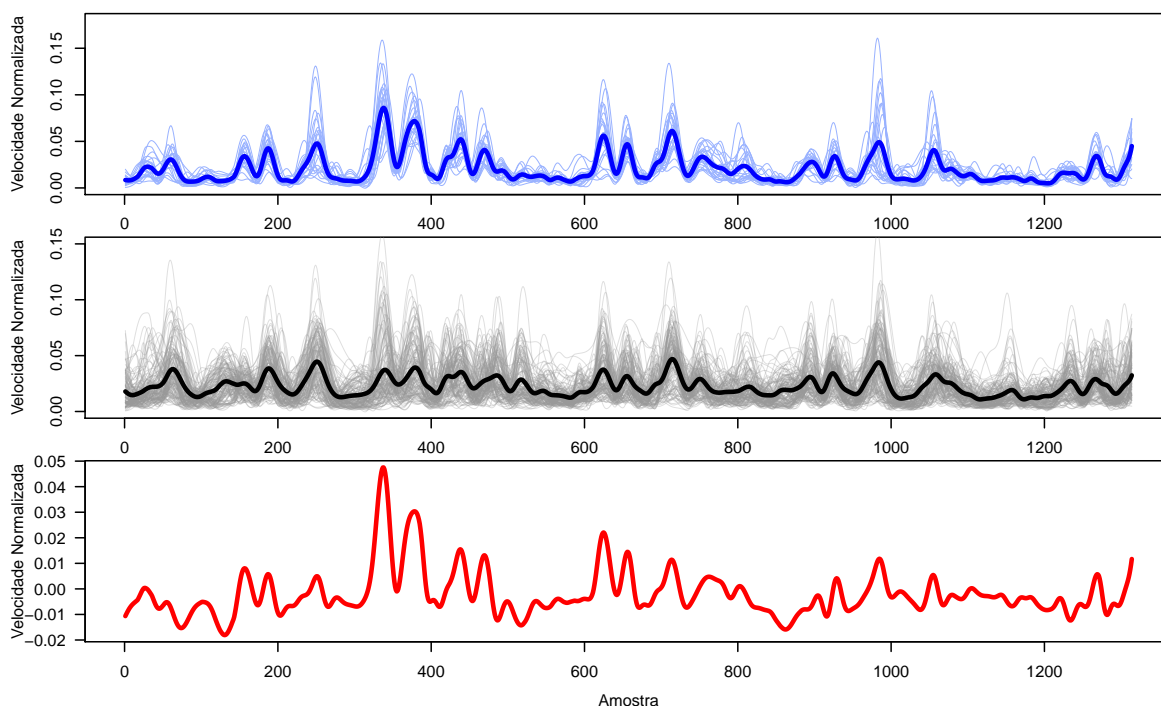


Figura 3.6: Painel superior: *Perfis de velocidade* das execuções gravadas pelo clarinetista **C4** e sua média superposta. Painel central: *Perfis de velocidade* de todas as gravações superpostos pela média global. Painel inferior: retro-projeção dos componentes LDs para o clarinetista **C4** no espaço original dos dados.

Após o processamento dos dados pela PCA, os autovetores retornados são ordenados de forma decrescente em relação a sua variância (autovalores), portanto, ao fazermos a retro-projeção destes vetores acabamos por selecionar apenas as regiões com maior variância entre as execuções. De fato, podemos observar que os valores da retro-projeção indicam a tendência de variação da média individual em relação à média global. Isto acontece devido à rotação executada pela LDA, a qual coloca em relevo as variações que melhor separam o grupo das execuções do clarinetista **C4** de todas as outras. Como exemplo, citamos a região compreendida entre as amostras 300 e 400. Nesta região, é possível identificar dois picos mais salientes no vetor retro-projetado. Se analisarmos a média dos *perfis de velocidade* deste músico, veremos que ele tende a realizar movimentos com maior velocidade nesta região em comparação com os demais músicos. Este padrão de movimentação poderia estar relacionado, por exemplo, a uma especificidade interpretativa, como uma articulação diferenciada das notas contidas nesta porção da música.

Como demonstrado na Figura 3.6, a retro-projeção evidencia os pontos (amostrados temporalmente) que possuem maior variância, em relação à média global, responsáveis pela discriminação entre os grupos sugeridos. Tendo isso em vista, na Figura 3.7, o vetor com os valores médios dos componentes LD para cada clarinetista é retro-projetado no espaço dimensional original e somado à média global, a fim de salientar estas regiões. A média global da população é representada na cor cinza, a média das execuções de cada grupo é representada na cor azul, e o resultado da retro-projeção, somada à média global, é mostrado em vermelho. A retro-projeção das regiões com maior variância permite a investigação das especificidades existentes nos *perfis de velocidade* de cada clarinetista. Novamente, é notável o fato de que as retro-projeções tendem a seguir a direção da variação das médias dos grupos, nos pontos mais importantes para a separação entre eles.

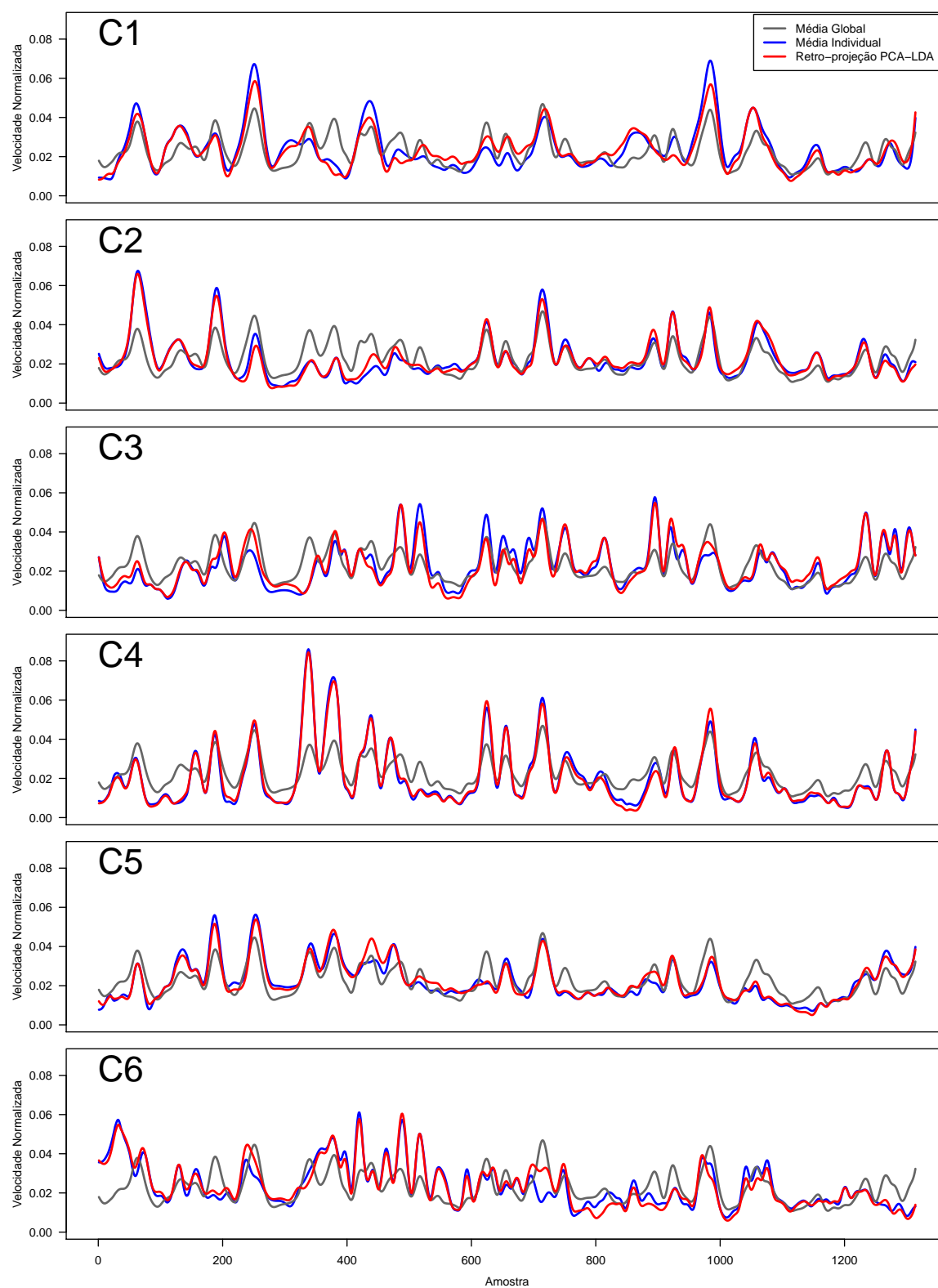


Figura 3.7: Retro-projeção dos componentes LD para o agrupamento por *seconds*, separadas por cada clarinetista (linhas vermelhas). A média da população (linhas cinzas) e as médias individuais (linhas azuis) também são mostradas.

3.1.2 Agrupamento por *primos*

Para este agrupamento, o cálculo recursivo LDA-PCA mostrou que, para uma discriminação sem erros, são necessários 86 PCs. Este valor é muito superior ao número de PCs necessários para discriminação do agrupamento por *secondos*, que foi de 10 PCs. O número de classificações erradas é demonstrado na Figura 3.8, e o resultado da classificação na Figura 3.9.

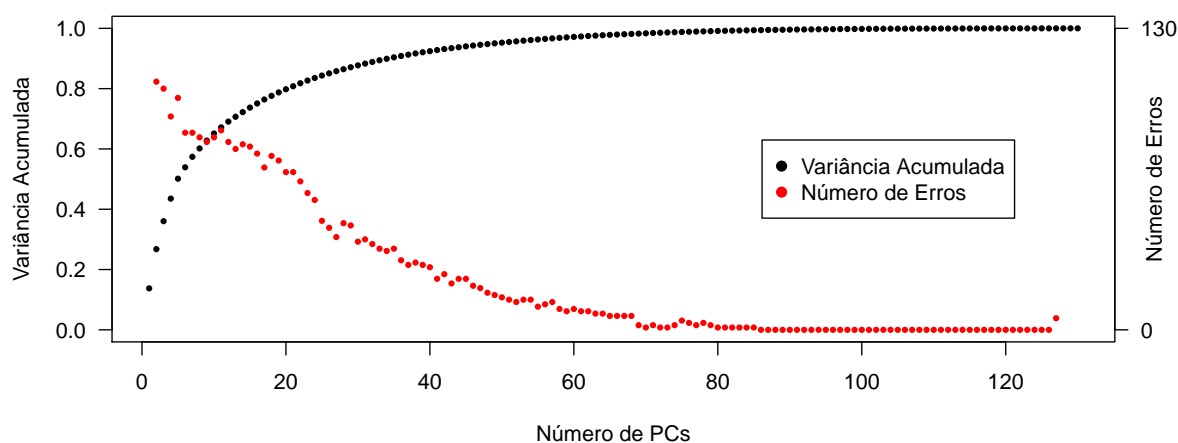


Figura 3.8: Número de erros retornado pela combinação PCA-LDA para o agrupamento por *primos*. A variância acumulada também é representada.

Retro-projeção dos componentes LD - Agrupamento por *Primos*

Repetimos o procedimento de retro-projeção dos vetores com os valores médios dos componentes LD para os clarinetistas *primos*. A influência de cada *primo* pode ser visualizada como a diferença entre as médias dos *secondos* e a média global, como exibido na Figura 3.10. Devido ao número elevado de PCs utilizados, uma menor variação em torno da média global é esperada. Neste caso, multiplicamos os vetores retro-projetados por um fator. Este é certamente um procedimento artificial, mas que permite uma melhor visualização dos efeitos decorrentes da influência de cada *primo* nos *perfis de velocidade* dos outros músicos.

O efeito mais notável neste caso é o perfil ruidoso das retro-projeções. Isto era esperado, considerando o número elevado de PCs necessários para a discriminação. Por outro lado, não parece existir nenhum efeito relevante nas retro-projeções que possa ser associado a uma tendência de movimentação dos *secon-*

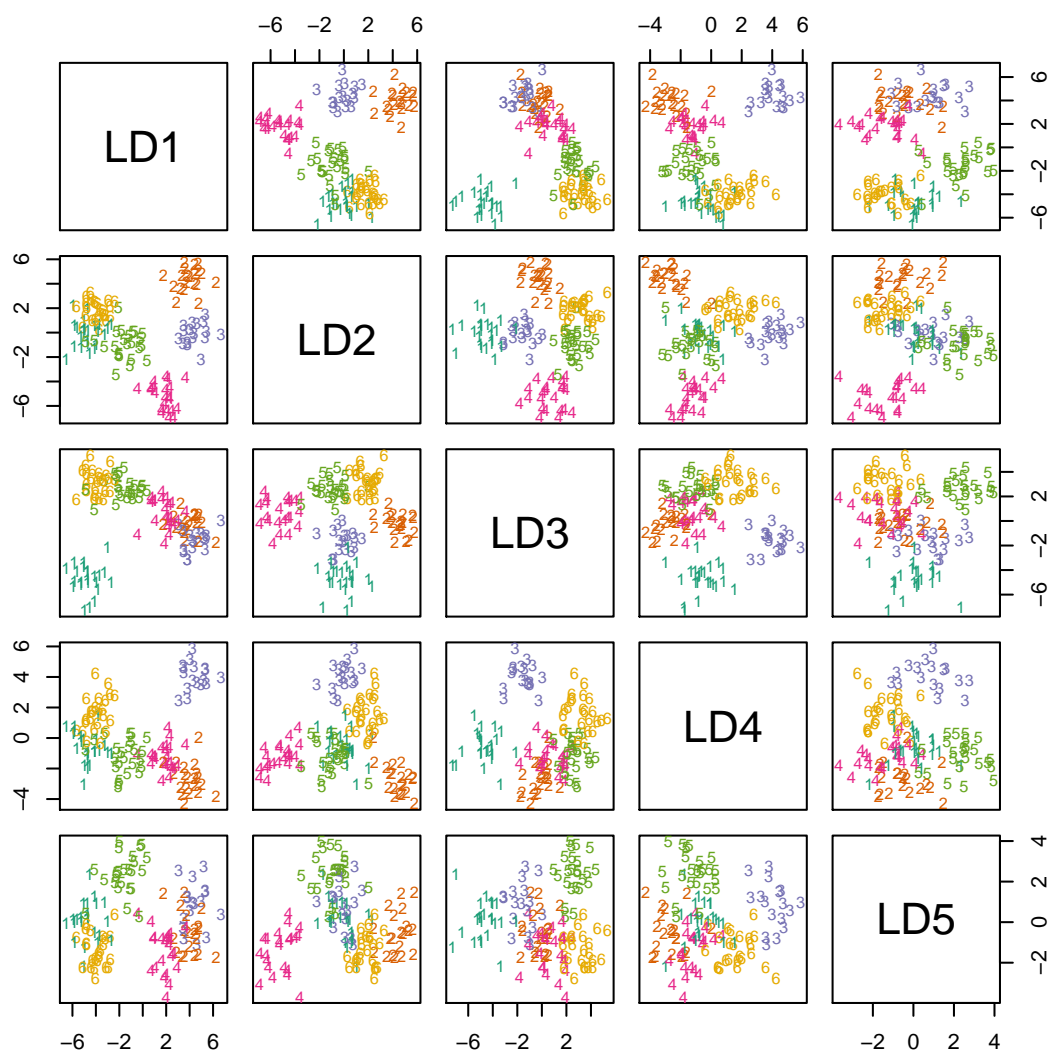


Figura 3.9: Resultados da LDA para o agrupamento por *primos*, após redução do espaço dimensional pela PCA.

dos quando estes acompanham diferentes *primos*. Esse resultado não é surpreendente pois, como argumentamos anteriormente, cada músico parece possuir um perfil de movimentação bastante recorrente e idiossincrático.

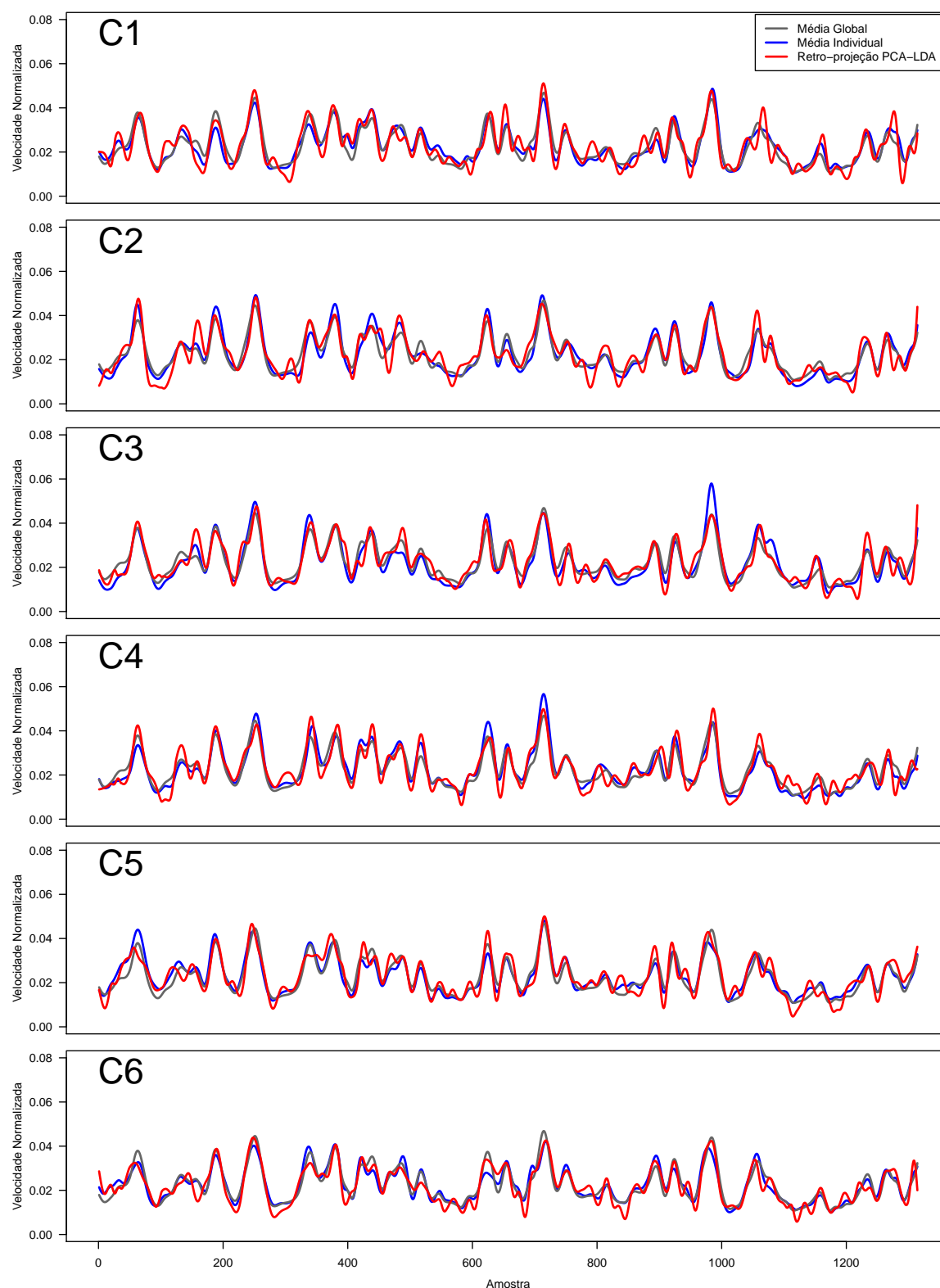


Figura 3.10: Retro-projeção dos componentes LD para o agrupamento por *primos*, separadas por cada clarinetista (linhas vermelhas). A média da população (linhas cinzas) e as médias individuais (linhas azuis) também são mostradas.

3.2 Discriminação dos dados acústicos

Como demonstrado na sessão anterior, o uso combinado de PCA e LDA para a discriminação dos dados cinemáticos em classes de clarinetistas *secondos* revelou a existência de um padrão próprio de movimentação de cada músico, o que poderia ser visto como uma “assinatura” gestual de cada intérprete. Esta “assinatura” parece sofrer pouca influência das diferentes condições testadas, isto é, não parece haver modificação significativa nos *perfis de velocidade* independente de qual *primo* clarinetista está sendo acompanhado. A retro-projeção dos discriminantes lineares computados permitiu ainda a determinação dos períodos temporais mais relevantes para a discriminação dos agrupamentos propostos. Nesta sessão, aplicamos a LDA aos dados acústicos das execuções analisadas, de forma a testar o poder de discriminância dos diferentes descritores.

3.2.1 Considerações iniciais

Considerando que os descritores utilizados nesta análise são escalares relacionados a cada nota musical, estas foram consideradas como coordenadas do espaço na análise, ao invés de amostras temporalmente equidistantes, como foi no caso do dados gestuais. Assim, a LDA foi aplicada aos dados acústicos com uma média de 34 notas (coordenadas) por observação, já que o excerto testado possui 41 notas, das quais, em média 17% foram excluídas, dependendo do descritor. Esta análise utilizou apenas as gravações da segunda sessão, (segundos clarinetistas acompanhando os primeiros) e propôs os mesmos grupos sugeridos para os dados gestuais, *primos* e *secondos*. Assim, o agrupamento por *primos* busca identificar nas gravações dos segundos clarinetista, qual *primo* que ele acompanhava, a partir do efeito que ele induziria na interpretação do clarinetista *secondo* que o acompanhava.

3.2.2 Discriminação

No caso do agrupamento por *secondos* clarinetistas, apenas o descritor *Loudness* apresenta uma separação completa das classes. Os descritores *Índice de Legato* e *IOI Normalizado* mostram uma separação razoável, com respectivamente 98.52% e 97.05% de classificações corretas. Os descritores *Índice de Articulação* e

Logaritmo do Tempo de Ataque possuem uma separação um pouco menos efetiva, com respectivamente 90.44% e 86.76% de classificações corretas. Estes resultados preliminares mostram que é possível obter uma separação aceitável para este agrupamento. A Figura 3.11 apresenta uma representação gráfica da densidade da distribuição dos valores previstos para cada classe em cada um dos 5 LDs. As densidades foram computadas utilizando a técnica de estimação da densidade Kernel implementada pelo pacote (VENABLES; RIPLEY, 2002). Os valores foram calculados para o descritor *Loudness*.

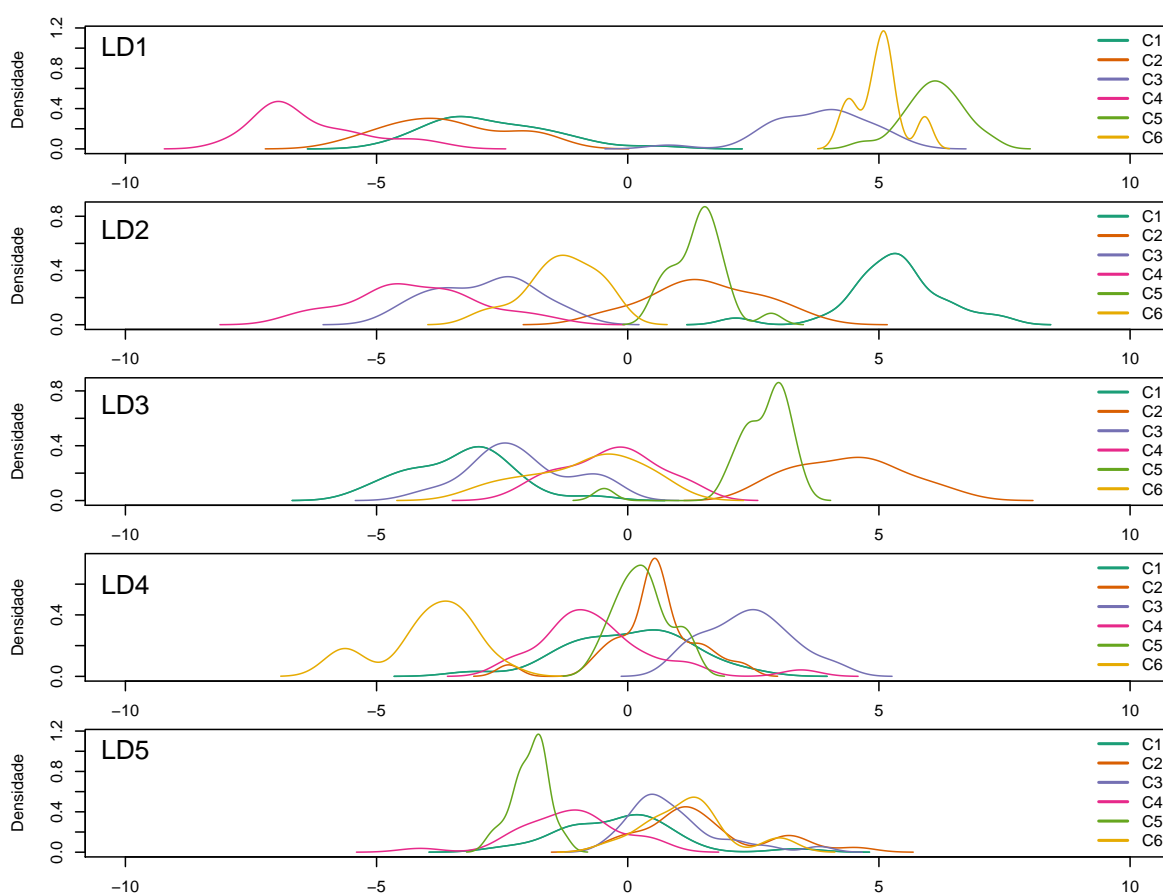


Figura 3.11: Gráfico da distribuição dos valores previstos para cada classe nos 5 LDs. Os valores foram calculados para o descritor *Loudness*, para o agrupamento por *seconds*. Curvas muito superpostas indicam pouca separação entre as classes.

Para o agrupamento por *primos*, o descritor *IOI Normalizado* obteve o melhor resultado com 97.05% de classificações corretas. Os descritores *Índice de Articulação* e *Loudness* apresentaram performance similar com 72.79% e 73.52% de classificações corretas respectivamente. O *Índice de Legato* obteve 69.85% e o *Logaritmo*

do *Tempo de Ataque* apenas 50% de separação. A Figura 3.12 mostra um gráfico da distribuição dos valores previstos para cada classe nos 5 LDs. Os valores foram calculados para o descritor *IOI Normalizado*. Estes resultados sugerem que o *IOI Normalizado* poderia ser utilizado para prever qual o primeiro clarinetista está sendo acompanhado nas gravações. No entanto, estes resultados refletem apenas a capacidade de separação dos dados originais, isto porque os dados utilizados na classificação foram os mesmos utilizados para treinar o modelo. Os resultados não são portanto suficientes para argumentar sobre o poder de classificação de cada descritor.

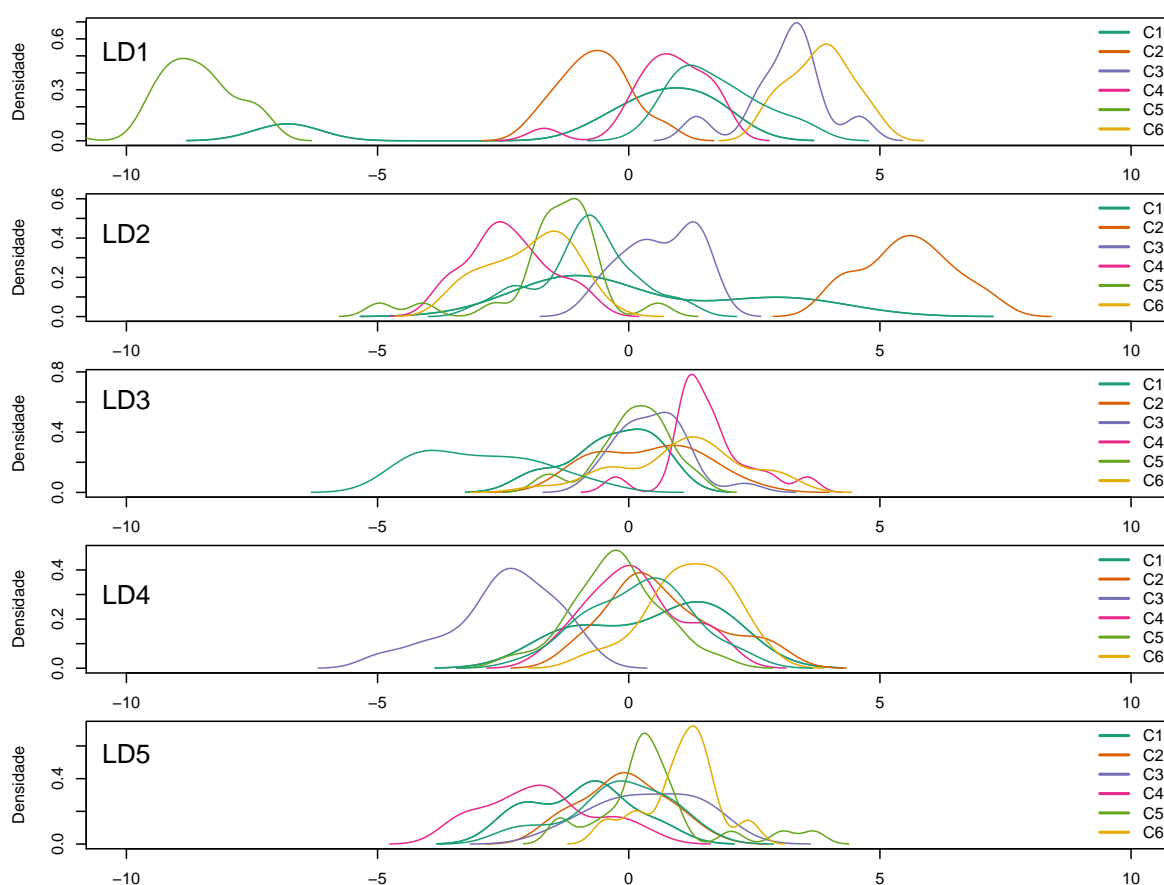


Figura 3.12: Gráfico da distribuição dos valores previstos para cada classe nos 5 LDs. Os valores foram calculados para o descritor *IOI normalizado*, para o agrupamento por *primos*. Curvas muito superpostas indicam pouca separação entre as classes.

3.3 Determinação de regiões significantes por retro-projeção

A complexidade das condições experimentais propostas resultou em uma quantidade substancial de dados coletados. Parte destes dados, tanto acústicos quanto cinemáticos, possuem informações redundantes ou de pouco valor analítico. A redução do volume de dados é aconselhável pois ela permite uma análise mais conveniente e uma exploração mais rápida dos dados. A retro-projeção dos discriminantes lineares (LDs) no espaço original dos dados pode, a princípio, permitir a identificação das variáveis originais que mais contribuem para a separação das classes. A Sessão 2.7.3 demonstrou que este procedimento pode ser utilizado como métrica para quantificar a contribuição de cada variável para a classificação realizada. Este procedimento foi aplicado ao agrupamento por *seconds* clarinetistas, de forma a encontrar as regiões mais significantes para caracterizar os grupos.

Obviamente algumas passagens de um excerto musical possuem um peso maior para a caracterização do estilo interpretativo de um músico. Estas podem ser o final de uma frase onde sempre é realizado um *rallentando*, ou um *crescendo* em uma sequência de notas repetidas, ou ainda uma forma específica de resolver um trinado. Músicos diferentes possuem formas diferentes de lidar com o material musical, e são estas diferenças que permitem a identificação do estilo de cada um. A concentração da análise em passagens mais significativas do excerto pode facilitar a identificação das características individuais de cada músico, sua “assinatura interpretativa”.

A Figura 3.13 apresenta a superposição dos valores de significância calculados para os *perfis de velocidade* e para cinco descritores acústicos, são eles: *Índice de Articulação*, *IOI Normalizado*, *Índice de Legato*, *Logaritmo do Tempo de Ataque* e *Loudness*. Podemos verificar que algumas regiões significantes para o domínio gestual coincidem com as regiões significantes para o domínio acústico. Na Figura 3.13 as barras verticais indicam os valores de significância dos descritores acústicos para cada uma das 41 notas do excerto. A curva contínua representa os valores de significância para cada ponto amostrado dos *perfis de velocidade*.

A relação entre as regiões mais significativas para os dados acústicos e os dados cinemáticos sugere que os músicos tendem a se mover de forma diferenciada

em regiões críticas do excerto musical. A região compreendida entre as notas 10 e 15 contém grandes articulações, onde podemos observar valores altos de significância para o descritor *Índice de articulação*. Este resultado se repete para as notas 30 e 31, que possuem articulação similar às notas 10 e 11. Nestas regiões observamos valores igualmente altos de significância para os dados gestuais, o que pode indicar uma possível relação entre os dois parâmetros. Observamos ainda que o início e o final do excerto possuem maior significância nos dados cinemáticos, o que indica que os músicos tendem a fazer movimentos mais específicos no início e o final da frase.

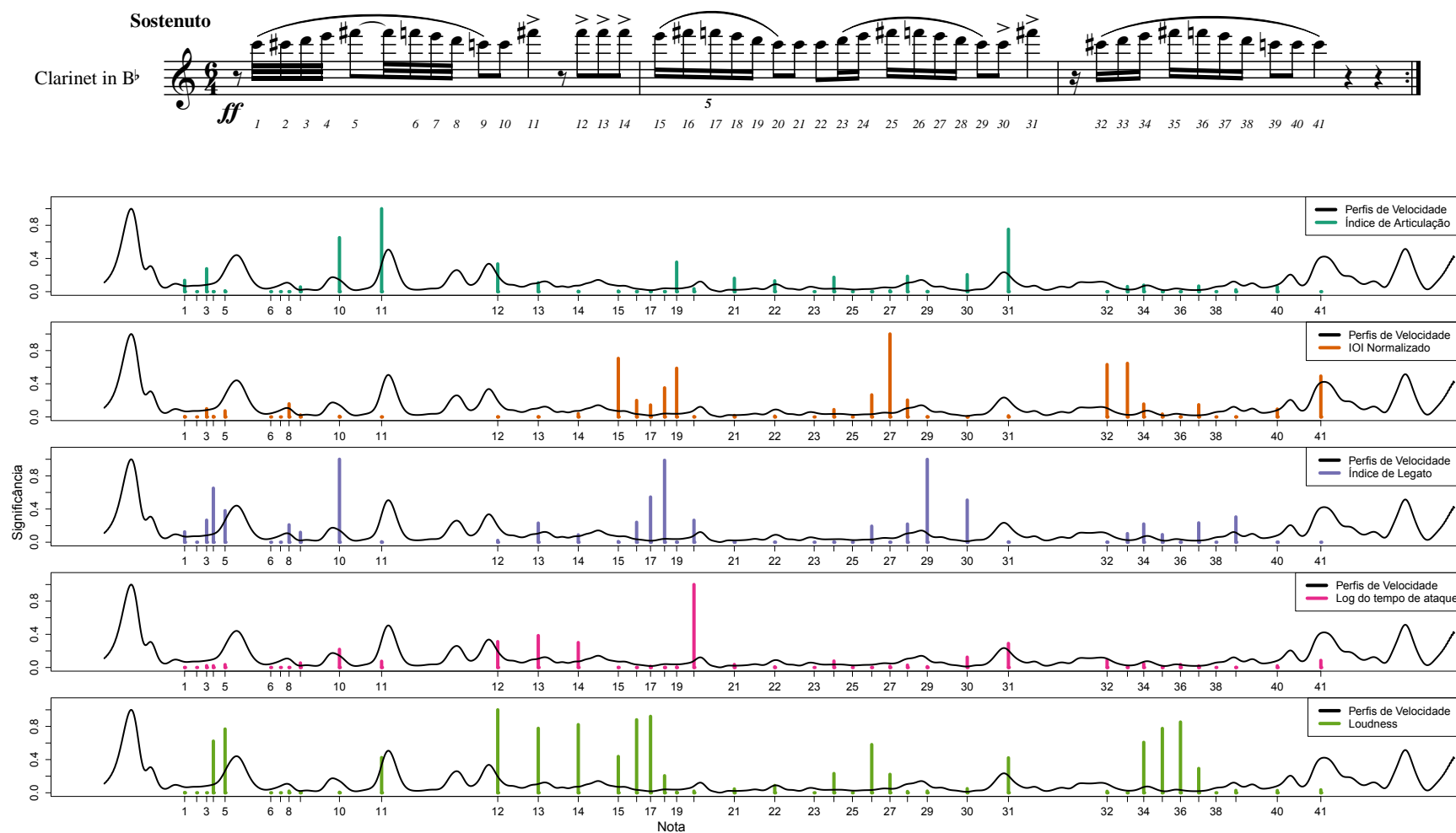


Figura 3.13: Superposição dos valores de significância calculados para os *perfis de velocidade* e para cinco descritores acústicos: Índice de Articulação, IOI Normalizado, Índice de Legato, Logaritmo do Tempo de Ataque e *Loudness*. Regiões com valores altos possuem maior importância para a discriminação entre os grupos.

3.4 Efeito *self-other*

3.4.1 Dados acústicos

Para realizar a análise do efeito *self-other* utilizando os dados acústicos, distribuimos os conjuntos seguindo o modelo proposto por Keller, Knoblich e Repp (2007), onde foi analisada a sincronia em duos de piano. No estudo realizado por Keller, Knoblich e Repp (2007) os resultados foram baseados na análise das diferenças temporais (assincronias) calculadas entre pianistas *primos* e *secondos*. Apenas um descritor (tempo de início das notas) foi utilizado. Os resultados reportados indicam que os pianistas conseguiram uma maior sincronização ao acompanhar a si próprios do que ao acompanhar os outros. A hipótese levantada pelos autores supõe que a sincronia é obtida por cada intérprete ao simular internamente as ações dos outros membros do conjunto, baseando-se inicialmente em como eles mesmos realizariam aquele excerto. Partindo desta hipótese, estendemos este estudo com a utilização de um maior número de descritores acústicos, tendo como um dos objetivos inferir sobre a capacidade de acompanhamento dos músicos sob diferentes parâmetros musicais, além do supracitado efeito *self-other*.

A distância entre as execuções foi definida como o módulo das diferenças dos valores de descritores atribuídos à cada nota, calculada separadamente para cada descritor. As execuções realizadas ao acompanhar sua própria gravação *solo* compõem o grupo *self* enquanto as execuções realizadas ao acompanhar as gravações *solo* dos outros compõem o grupo *other*. A normalidade das distribuições dos grupos de dados não foi confirmada (teste Shapiro-Wilk), por isso optamos por utilizar um teste não-paramétrico (Wilcoxon–Mann–Whitney) para a comparação entre as distribuições dos grupos *self* e *other*, apresentado na Figura 3.14 e na tabela 3.1.

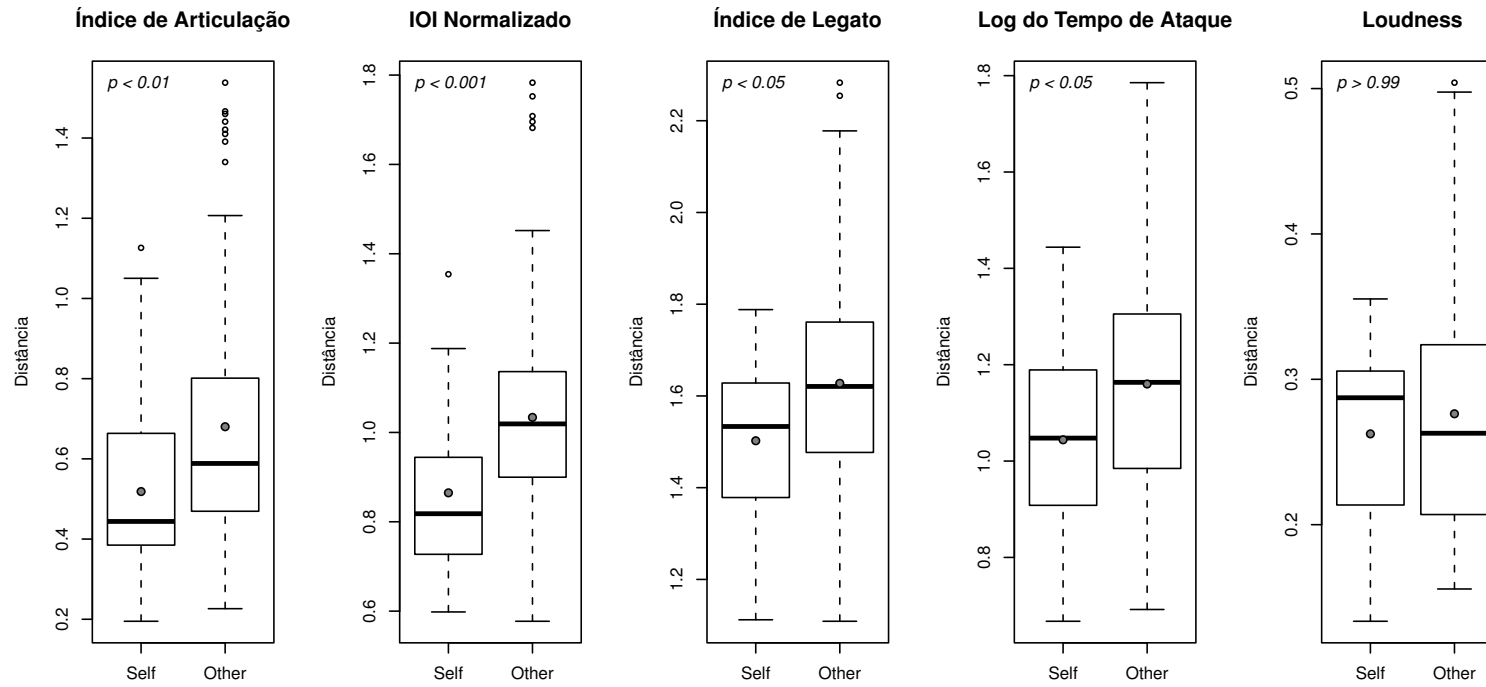


Figura 3.14: Boxplot dos descritores acústicos analisados para o efeito *self-other*. Valor- p calculado com o teste Wilcoxon–Mann–Whitney. As medianas são indicadas pela linha horizontal, as médias por círculos preenchidos e os *outliers* por círculos vazios.

Tabela 3.1: Diferenças dos descritores acústicos analisados para o efeito *self-other*. Valor-*p* calculado com o teste Wilcoxon–Mann–Whitney.

<i>Descritor</i>	<i>Self</i>		<i>Other</i>		W	Valor- <i>p</i>
	<i>Média</i>	<i>d.p.</i>	<i>Média</i>	<i>d.p.</i>		
<i>Índice de Articulação</i>	0.518	0.238	0.68	0.307	816	< 0.01
<i>IOI Normalizado</i>	0.865	0.170	1.03	0.236	661	< 0.001
<i>Índice de Legato</i>	1.50	0.173	1.63	0.235	899	< 0.05
<i>Logaritmo do Tempo de Ataque</i>	1.04	0.183	1.16	0.229	890	< 0.05
<i>Loudness</i>	0.26	0.071	0.27	0.088	1274	> 0.99

Quatro dos cinco descritores acústicos testados apresentaram diferenças significativas entre as execuções *self* e *other*, sendo que o resultado mais significativo foi observado para o descritor *IOI Normalizado* ($p < 0.001$). Como detalhado na Sessão 3.2, este descritor depende do cálculo das posições temporais (*onsets*) das notas, ele é portanto relacionado à sincronia temporal entre as notas, observada nos estudos anteriores acima mencionados, Keller, Knoblich e Repp (2007) e Loureiro et al. (2012). O descritor *Loudness* não apresentou diferenças significativas. De forma geral, estes resultados indicam que os clarinetistas participantes do experimento tendem a seguir com maior eficiência suas próprias execuções, não só em termos rítmicos, mas também em outros parâmetros musicais.

3.4.2 Dados cinemáticos

Na Sessão 3.3 foi demonstrado que algumas regiões temporais possuem valores baixos de significância para a discriminação entre os intérpretes. Estas regiões são de pouca utilidade para a caracterização de uma “assinatura gestual” dos músicos. A seleção das regiões mais significantes permite a eliminação dos valores menos relevantes, contribuindo assim para uma interpretação mais clara dos dados. Portanto, para a análise do efeito *self-other* nos dados cinemáticos, selecionamos as regiões com valores mais altos de significância para os dados gestuais, que correspondem ao início (até a nota 15) e ao final (após a nota 30) do excerto.

A hipótese que motiva a análise do efeito *self-other* nos dados cinemáticos

prevê que, ao tocar em duo, o padrão gestual do segundo clarinetista se desloca em direção ao padrão gestual do primeiro clarinetista. No caso em que os clarinetistas acompanham suas próprias gravações, o seu padrão gestual não sofreria grandes alterações. Para verificar a ocorrência deste efeito é necessário a definição de uma “referência gestual” de cada músico. Para tanto, utilizamos o *perfil de velocidade* extraído da execução realizada como clarinetista *primo*, gravada na primeira sessão do experimento.

A similaridade entre os *perfis de velocidade* é calculada pela distância euclidiana entre eles. Considerando dois *perfis de velocidade* Q e C , ambos com número de amostras n a distância entre eles é dada pela raiz quadrada da soma dos quadrados das diferenças entre cada par de pontos amostrados:

$$D(Q, C) = \sqrt{\sum_{i=1}^n (q_i - c_i)^2} \quad (3.1)$$

A distância entre as curvas é calculada após cada uma das curvas ser normalizada para média 0 e desvio padrão 1 (ou *z-score*):

$$z = \frac{x - \mu}{\sigma} \quad (3.2)$$

onde μ é a média do *perfil de velocidade* e σ o seu desvio padrão. Este procedimento é necessário para permitir a comparação dos valores individuais de diferentes populações.

Quando um clarinetista acompanha um outro, não podemos esperar que o seu perfil gestual se adeque completamente ao perfil gestual do outro. No entanto, algum tipo de adaptação é possível, mesmo sem contato visual entre eles durante os experimentos. Para facilitar a abordagem deste problema, foram realizados dois testes, um para testar se os clarinetistas se aproximam mais da sua referência ao acompanhar suas próprias gravações, e o outro para testar se os clarinetistas se aproximam mais da referência dos outros ao acompanhar as gravações deles.

Considerando S como a “referência gestual” de um dado clarinetista, O como a “referência gestual” do outro clarinetista, s como as execuções realizadas ao acompanhar sua própria referência e o como as execuções realizadas ao acompanhar a referência do outro.

I) Primeiramente testamos a hipótese

$$\begin{aligned} H_0 &: D(S,s) = D(S,o) \\ H_1 &: D(S,s) < D(S,o) \end{aligned} \quad (3.3)$$

onde a hipótese alternativa implica que as execuções realizadas ao acompanhar a si próprio s são mais próximas da sua própria referência S do que as execuções realizadas ao acompanhar um outro o .

II) Em seguida testamos a hipótese

$$\begin{aligned} H_0 &: D(O,s) = D(O,o) \\ H_1 &: D(O,s) > D(O,o) \end{aligned} \quad (3.4)$$

onde a hipótese alternativa implica que as execuções realizadas ao acompanhar o outro o são mais próximas da referência do outro O do que as execuções realizadas ao acompanhar a si próprio s . Uma representação esquemática do procedimento realizado é apresentada na Figura 3.15.

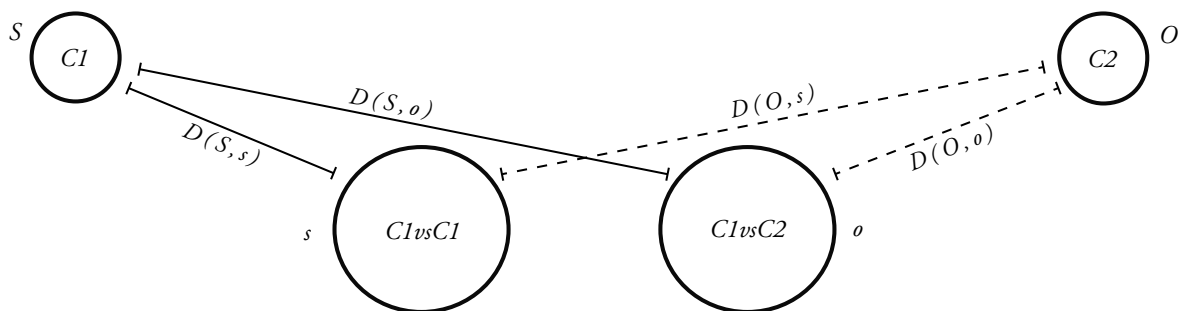


Figura 3.15: Representação esquemática do procedimento empregado no cálculo das distâncias *self-other*.

Para verificarmos a validade dos resultados, optamos por uma técnica não-paramétrica, o teste Wilcoxon–Mann–Whitney para amostras independentes (WOLFE; HOLLANDER, 1973). Este teste permite estimar o deslocamento de uma população para a esquerda ou para a direita em comparação com o outra. O número reduzido de amostras distribuídos para alguns grupos torna inviável a comprovação da normalidade. Esta técnica apresenta uma solução razoável para os testes propostos pois seus pressupostos formais são a independência entre as variáveis que devem ter sido medidas em uma escala ordinal e que suas distribuições sejam similares.

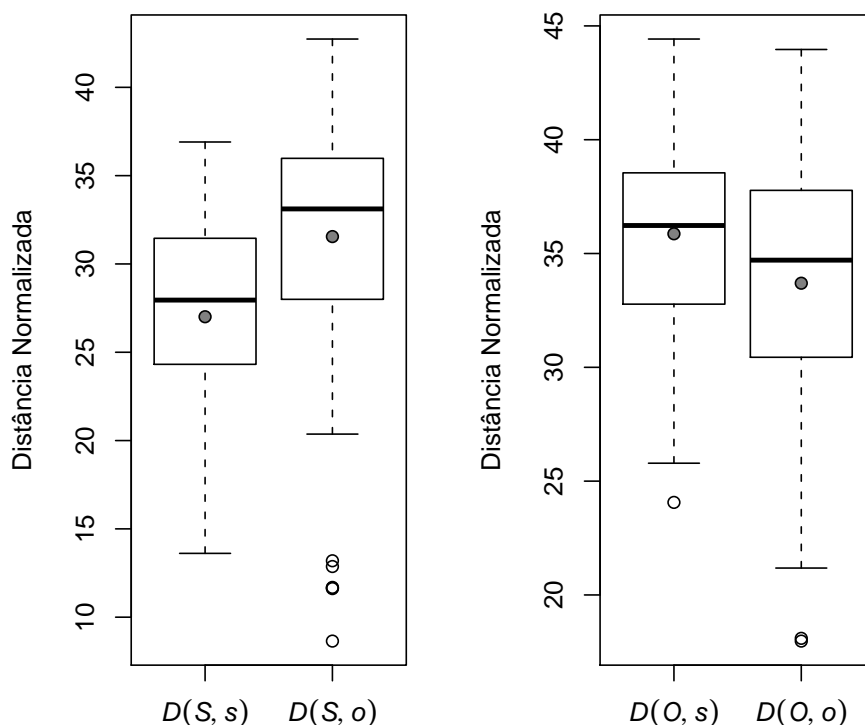


Figura 3.16: Comparação entre as distâncias dos conjuntos de dados $D(S, s)$ e $D(S, o)$ (painel à esquerda) e entre os conjuntos de dados $D(O, s)$ e $D(O, o)$ (painel à direita). As medianas são indicadas pela linha horizontal, as médias por círculos preenchidos e os *outliers* por círculos vazios.

Os resultados do primeiro teste (painel esquerdo da Figura 3.16) demonstraram que os valores das distâncias $D(S, s)$, com $M = 27.01$ e $d.p. = 5.80$ são significativamente menores que as distâncias $D(S, o)$, com $M = 31.550$ e $d.p. = 7.01$ ($p < .001$). Este resultado sugere que em média o padrão gestual dos músicos testados se distancia mais da sua própria “referência gestual” quando eles acompanham outros músicos do que quando eles acompanham a si mesmos. Pode-se argumentar que este resultado aponta para uma degradação do padrão gestual original, causada pela dificuldade técnica em realizar a tarefa proposta (seguir um músico desconhecido). Esta hipótese foi apontada por Wanderley et al. (2005) ao analisar os efeitos da performance de clarinetistas em passagens orquestrais tecnicamente avançadas. Por isso é necessário verificar se esse deslocamento ocorre em direção

ao padrão gestual dos outros, e não um deslocamento aleatório, confirmamos isto com aplicação do segundo teste.

Os resultados do segundo teste (painel direito da Figura 3.16) apontam que os valores das distâncias $D(O,s)$, com $M = 35.86$ e $d.p. = 3.91$ são significativamente maiores que as distâncias $D(O,o)$, com $M = 33.69$ e $d.p. = 5.64$ ($p < .01$). Este resultado aponta que em média o padrão gestual dos músicos testados se desloca em direção à “referência gestual” dos músicos acompanhados. Como afirmado anteriormente, durante os experimentos não houve contato visual ou qualquer outro tipo de interação entre os intérpretes que não fosse unicamente comunicada através do sinal de áudio. No entanto, foi possível verificar a tendência a um deslocamento do padrão gestual dos *secondos* em direção aos *primos*.

3.5 Avaliação da discriminação de performances

Assim como apontado na Sessão 3.1, cada um dos agrupamentos propostos foi dividido em seis classes, uma para cada clarinetista participante dos experimentos. O objetivo inicial desta análise foi testar a discriminação dos agrupamentos sugeridos e criar um modelo de classificação utilizando apenas dados gestuais. Posteriormente, esta análise foi também estendida aos dados acústicos. Com isso foi possível testar o poder de discriminação dos diferentes parâmetros acústicos utilizados nas gravações da segunda sessão, isto é, os casos onde os *secondos* clarinetistas seguem os *primos*.

A verificação de algoritmos de classificação é necessária para avaliar a real capacidade do modelo, isto é, confirmar se o modelo é capaz de classificar corretamente observações que não foram usadas no treinamento (DUDA; HART, 1996, cap. 9, p. 32). De acordo com Stone (1974), uma das técnicas de validação mais antigas e utilizadas é a chamada “validação-cruzada”. Esta técnica consiste em dividir os dados disponíveis aleatoriamente em conjuntos de treinamento e teste. O primeiro deles é utilizado para ajustar os parâmetros do modelo classificador, e o segundo é utilizado para o cálculo do erro de classificação. Ainda segundo Duda e Hart (1996, cap. 9, p. 33), é essencial que o grupo de teste não inclua observações utilizadas para treinar os parâmetros do classificador. Isto representa um problema para conjuntos de dados com número reduzido de observações.

Para contornar este problema são utilizadas técnicas que maximizam o número de iterações entre os dados particionados. Um dos procedimentos mais comuns é o chamado k -fold, onde o conjunto dos dados disponíveis \mathcal{D} é dividido aleatoriamente em k conjuntos de tamanho n/k , onde n é igual ao número de observações contidas em \mathcal{D} . O modelo classificador é então treinado k vezes, cada uma das vezes com uma combinação diferente de $k - 1$ partições e finalmente testados na partição restante. A performance estimada do classificador é calculada simplesmente pela média dos k erros de cada iteração Duda e Hart (1996, cap. 9, p. 35).

Se considerarmos $k = n$ temos um caso especial de K -fold conhecido como *leave-one-out*. Neste caso, a precisão do algoritmo de classificação é estimada treinando o classificador n vezes separadamente, utilizando em cada iteração o conjunto de treino \mathcal{D} , que teve uma única observação de cada grupo excluída. Cada treino resultante das diferentes iterações é testado contra estas observações que foram excluídas. Assim como no K -fold a estimativa da precisão é igual a média de todos os K erros (DUDA; HART, 1996, cap. 9, p. 39). Esta técnica faz bom uso dos dados disponíveis na medida em que cada observação no conjunto de dados é usada tanto no grupo de treinamento quanto no grupo de teste, portanto este foi o procedimento escolhido para a validação das classificações realizadas neste estudo.

3.5.1 Agrupamento por *SECONDOS*

Dados Cinemáticos

A validação-cruzada mostrou uma taxa média de discriminação de 95,40 % (d.p. = 5,96) com o uso de apenas 10 Componentes Principais para o agrupamento por *secondos* dos dados cinemáticos, como mostra a tabela 3.2. Este resultado evidencia a existência de características gestuais próprias à cada indivíduo testado, o que poderíamos chamar de uma “assinatura gestual” dos intérpretes. Estas características são fortemente recorrentes entre as execuções dos intérpretes e não parecem sofrer efeitos de diferentes condições de execução, isto é, quando estes intérpretes acompanham outros músicos seu padrão de movimentação individual se mantém na maior parte inalterado.

A tabela 3.2 mostra ainda que os valores de classificações corretas variam entre

Tabela 3.2: Resultado do teste de validação cruzada para a classificação dos dados cinemáticos. Agrupamento por clarinetistas *secondos*.

	Clarinetista						Class. Corretas (%)	
	C1	C2	C3	C4	C5	C6	M	d.p.
<i>N</i>	24	22	24	24	21	15	-	-
<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
<i>Class. Corretas (%)</i>	100	90.90	95.83	100	85.71	100	95.40	5.96

os diferentes clarinetistas, três deles apresentando 100 % de acerto. Este resultado sugere que a hipótese de “assinatura gestual” do intérprete pode se dar em duas dimensões: (1) músicos com menor índice de classificações corretas variam mais os seus gestos que os outros, ou seja, se afastam de seus padrões com mais facilidade; (2) músicos distintos podem apresentar similaridade entre seus padrões gestuais, o que dificultaria a discriminação entre eles.

Dados Acústicos

Como mostra a Tabela 3.3, o teste de validação-cruzada para a classificação dos dados acústicos, assim como para os dados cinemáticos, apresentou taxas elevadas de acertos para todos os descritores analisados, variando de 70,91 % (d.p. = 10,19) para o descritor *Logaritmo do tempo de ataque*, a 97,06 % (d.p. = 2,28) para o descritor *Loudness*. Estes valores elevados de classificações corretas sugerem a existência de uma “assinatura interpretativa” para cada músico, ou seja, cada um possui um padrão ou estilo interpretativo próprio, sugerindo que cada músico possua um planejamento próprio de como tocar este excerto. Além disso, o alto índice de discriminação observado em pelo menos 5 descritores, indica que o planejamento interpretativo de cada músico se manifesta a partir de manipulações do material acústico em diferentes dimensões, tais como andamento, dinâmica, agógica, articulação e timbre, e em graus distintos para cada dimensão.

Tabela 3.3: Resultado do teste de validação cruzada para a classificação dos dados acústicos.
Agrupamento por clarinetistas *secondos*.

<i>Descritor Acústico</i>		<i>Clarinetista</i>						<i>Class. Corretas (%)</i>	
		C1	C2	C3	C4	C5	C6	<i>M</i>	<i>d.p.</i>
<i>Índice de Articulação</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	58.33	77.27	79.16	54.16	90.47	80	73.23	14.00
<i>IOI Normalizado</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	87.50	59.09	83.33	66.66	85.71	100	80.38	14.92
<i>Índice de Legato</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	91.66	77.27	87.50	91.66	90.47	93.33	88.65	5.90
<i>Logaritmo do Tempo de Ataque</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	84.00	78.26	56.00	64.00	68.18	75.00	70.91	10.19
<i>Loudness</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	95.83	95.45	100	95.83	95.24	100	97.06	2.28

3.5.2 Agrupamento por *PRIMOS*

Dados Cinemáticos

O agrupamento por *primos* teve como objetivo principal testar o efeito das diferentes condições interpretativas impostas por cada intérprete enquanto tocando como *secondo*. Valores altos de classificações corretas para este agrupamento indicariam que os clarinetistas *secondos* adaptariam seu padrão gestual ao padrão gestual dos *primos*, o que seria uma evidência de um acoplamento entre os dois modos sensoriais, já que não houve contato visual entre os músicos durante as gravações.

Os resultados da validação cruzada para a classificação deste agrupamento apresentaram baixos valores de acerto, variando de 12,50 % a 35,00 %, como mostra a Tabela 3.4. Este resultado sugere que os clarinetistas *secondo* não se aproximaram do padrão gestual dos clarinetistas *primo* durante a performance em duo, mantendo seu próprio padrão, como mostrado na classificação do agrupamento em *secondos*. Este resultado era esperado, já que o número de Componentes Principais necessários para uma separação completa dos grupos de *primos* (3.1) foi elevado (86 PCs), significando que a separação só foi possível quando o modelo foi ajustado para considerar pequenas variações entre os *perfis de velocidade*, que ocorrem em velocidades baixas.

Tabela 3.4: Resultado do teste de validação cruzada para a classificação dos dados cinemáticos. Agrupamento por clarinetistas *primos*.

	<i>Clarinetista</i>						<i>Class. Corretas (%)</i>	
	C1	C2	C3	C4	C5	C6	<i>M</i>	d.p.
<i>N</i>	24	22	24	24	21	15	-	-
<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
<i>Class. Corretas (%)</i>	13.04	35.00	29.41	22.73	20.83	12.50	22.25	8.89

Dados Acústicos

A Tabela 3.5 mostra que apenas o descritor *IOI Normalizado* apresentou valor elevado de acertos, 90.12 % (d.p. = 8.54), enquanto os outros descritores apresentaram índices que variaram entre 21.33 % e 36.98 %. Este resultado indica que é possível identificar qual o clarinetista *primo* está sendo acompanhado, a partir dos valores de *IOI Normalizado* das execuções dos *secondos*, mas não a partir dos valores dos outros descritores.

Lembrando que, na segunda sessão dos experimentos, a situação de acompanhamento de um líder em uníssono, demandou que os clarinetistas *secondos* se aproximem ao máximo das execuções dos clarinetistas *primos*, o que implicaria em se aproximar ao máximo das intenções interpretativas do clarinetista *primo* percebidas a partir de suas manipulações do material acústico que os diferentes descritores buscam evidenciar, este resultado sugere que, nesta situação, músicos buscam se concentrar mais nos aspectos rítmicos, para garantir a sincronia da performance.

Tabela 3.5: Resultado do teste de validação cruzada para a classificação dos dados acústicos.
Agrupamento por clarinetistas *primos*.

<i>Descritor Acústico</i>		<i>Clarinetista</i>						<i>Class. Corretas (%)</i>	
		C1	C2	C3	C4	C5	C6	<i>M</i>	<i>d.p.</i>
<i>Índice de Articulação</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	26.09	50.00	17.65	40.91	50.00	29.17	35.63	13.39
<i>IOI Normalizado</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	100	90.00	88.24	100	79.17	83.33	90.12	8.54
<i>Índice de Legato</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	34.78	40.00	17.65	13.64	37.50	33.33	29.48	11.04
<i>Logaritmo do Tempo de Ataque</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	21.74	10.00	11.76	13.64	37.50	33.33	21.33	11.70
<i>Loudness</i>	<i>N</i>	24	22	24	24	21	15	-	-
	<i>Prob. a Priori (%)</i>	18.46	16.92	18.46	18.46	16.15	11.53	-	-
	<i>Class. Corretas (%)</i>	39.13	35.00	23.53	40.91	50.00	33.33	36.98	8.80

4 *Discussão*

Neste estudo investigamos os padrões de sincronização utilizados na performance instrumental de *duos* de clarineta, utilizando informações acústicas e gestuais. Para tanto, foram analisados os movimentos corporais e os parâmetros acústicos da performance de seis clarinetistas em uma situação simulada de orquestra. No experimento realizado, os músicos assumiram o papel de segundos clarinetistas e foram instruídos a seguir, da melhor forma possível, o primeiro clarinetista. Durante os experimentos não houve qualquer interação entre os músicos que não fosse através do sinal de áudio. Neste estudo consideramos apenas os gestos não essenciais para a produção sonora, chamados “acompanhantes”.

O objetivo deste estudo foi investigar os padrões de sincronização utilizadas em grupos musicais, partindo de duas abordagens complementares, análise cinematográfica e análise de parâmetros acústicos. Esta investigação partiu da análise da sincronia entre músicos tocando em conjunto, apresentadas em estudos anteriores. Estes estudos, que se restringiram às manipulações temporais das notas, (instantes de inícios de cada nota), verificaram que músicos apresentam maior sincronia ao acompanhar suas próprias performances gravadas, se comparada com gravações de outros músicos, um efeito conhecido como *self-other*. Partindo destes resultados, este estudo propôs verificar se esta sincronização observada poderia ser refletida: (1) na movimentação dos músicos enquanto tocam; e (2) em outras dimensões das manipulações do material acústico, tais como intensidade, qualidade de articulação ou colorido sonoro.

Abordamos o problema em duas etapas, primeiramente foi testada a discriminação entre as execuções, sob dois diferentes agrupamentos, *primos* e *secondos* clarinetistas. Para isso foi utilizada uma combinação de PCA e LDA, que permitiu reduzir o número de pontos utilizados na análise ao mínimo necessário para uma separação entre as classes. O objetivo nesta etapa foi uma exploração inicial

dos dados que permitisse mostrar o quão recorrente são os padrões gestuais dos clarinetistas, e se estes padrões eram comuns a todos ou específicos de cada um. Os resultados para o agrupamento por *secondos* mostraram que os padrões gestuais dos músicos possuem uma forte recorrência entre as diferentes performances. Com isso foi possível constatar a existência de uma “assinatura” gestual de cada intérprete. Isto permitiu a separação das execuções dos clarinetistas utilizando apenas os dados gestuais.

Para comprovar o poder de separação deste agrupamento foi realizado um teste de validação cruzada (*leave-one-out*). Os resultados demonstraram uma taxa de 95,4% para as classificações corretas. O agrupamento por *primos* tinha como objetivo testar se os perfis de movimentação dos *secondos* sofriam perturbações enquanto estes acompanhavam diferentes *primos*. Os resultados mostraram a existência de pequenas variações em regiões críticas do excerto musical, mas estas não foram suficientes para a separação dos grupos e a consequente identificação dos *primos* que estavam sendo acompanhados. Os resultados do teste de validação cruzada apontaram baixas taxas de classificações corretas para este agrupamento (22.25%).

A discriminação LDA foi aplicada também aos dados acústicos, formalizados em cinco descritores distintos seguindo o mesmo padrão de agrupamento. Os resultados do teste de validação cruzada para o agrupamento por *secondos* demonstram que os músicos possuem uma “assinatura” musical que pode ser acessada pelos valores nota-a-nota dos descritores. Os resultados do teste de validação cruzada para o agrupamento por *primos* mostraram taxas de acertos que variaram entre uma média de 21.33% para o descritor *Logaritmo do Tempo de Ataque* e uma média de 90.12% para o descritor *IOI Normalizado*, sugerindo que músicos priorizam “seguir” certos parâmetros musicais ao acompanhar outro músico e são mais bem sucedidos na tarefa de sincronizar os início de notas.

Acreditamos que isto aponta para uma hierarquia dos parâmetros musicais manipulados neste excerto. Isto é, toda vez que os músicos testados eram confrontados com uma situação interpretativa diferente (um novo *primo* para acompanhar) eles precisavam adequar a sua performance. Com a impossibilidade de ajustar todos os parâmetros musicais necessários, é possível que eles tenham se concentrado naqueles que foram mais evidentes, neste caso, as variações temporais induzidas pelos *primos* indicadas pelo descritor *IOI Normalizado*. Ou seja,

o modelo de classificação criado para este descritor é capaz de identificar qual o *primo* clarinetista estava sendo seguido, tendo como informação apenas os dados extraídos da performance do *secondo* clarinetista. Este resultado não é surpreendente, visto que ao tocar em duetos, os músicos buscam a maior adequação possível dos parâmetros musicais indicados pelo líder. A sugestão de uma hierarquia entre os parâmetros musicais neste excerto abre caminho para questões futuras. Uma possibilidade seria investigar se este resultado é influenciado por outros fatores (complexidade técnica, estilo da peça, tessitura da frase, etc).

Uma das vantagens da LDA é que os discriminantes lineares podem ser retro-projetados no espaço original dos dados. O resultado deste procedimento pode ser utilizado para calcular a significância discriminatória das variáveis utilizadas na LDA. As variáveis utilizadas para a discriminação entre as classes foram discretizadas temporalmente. No caso dos dados gestuais, as variáveis correspondem aos valores de módulo da velocidade calculados das 1500 posições espaciais amostrados a uma taxa de 100 Hz. Para os dados acústicos, as variáveis correspondem a cada uma das 41 notas. Com isto foi possível identificar quais são as regiões temporais mais significantes para a discriminação entre os músicos. A superposição das retro-projeções dos discriminantes acústicos e dos discriminantes cinemáticos, permitiu a comparação entre as regiões mais significantes nos dois domínios, permitindo a exclusão das regiões com informação redundante ou pouco significativa. Estas regiões foram utilizadas para a análise realizada na etapa seguinte.

Finalmente, o estudo buscou verificar a ocorrência do efeito *self-other*, nos dados gestuais e acústicos, evidenciada em estudos anteriores, não apenas em performances musicais (KELLER, 2001; KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010; LOUREIRO et al., 2012), mas também em outras tarefas que envolvem ação-percepção, tais como o reconhecimento de caligrafia (KNOBLICH et al., 2002) e no lançamento de dardos (KNOBLICH; FLACH, 2001). Evidências da ocorrência deste efeito no domínio musical foram demonstradas em diferentes estudos (KELLER, 2001; KELLER; KNOBLICH; REPP, 2007; KELLER; APPEL, 2010). A hipótese mais aceita para explicar este efeito propõe que a sincronia entre os músicos é alcançada por cada intérprete simulando internamente as ações dos outros membros do conjunto, baseando-se inicialmente em como eles mesmos realizariam a performance daquele excerto. Portanto, ao acompanhar a si próprios, eles reconheceriam as ações musicais do clarinetista *primo* como auto-produzidas e se adequariam com maior facilidade a elas. O efeito *self-other* para os dados acús-

ticos foi testado para diferentes descritores e apresentou resultados condizentes com os relatados por Keller, Knoblich e Repp (2007), sugerindo que os músicos tendem a seguir com maior eficiência suas próprias execuções, não só em termos rítmicos, mas também em outras dimensões acústicas. Este resultado é significativo pois estende a análise realizada previamente em outros estudos.

Visando verificar o efeito *self-other* para os dados cinemáticos, buscamos demonstrar que, quando músicos acompanham outros músicos, seu padrão gestual se distancia daquele padrão apresentado quando toca sozinho (execuções *solo*). Os resultados sugeriram esta tendência, o que poderia sinalizar uma maior dificuldade em cumprir a tarefa proposta, assim como apontado por Wanderley et al. (2005), onde a indicação para tocar com diferentes níveis de expressividade provocou uma variação na amplitude dos movimentos dos músicos. Neste caso os músicos teriam simplesmente mais dificuldade em acompanhar os outros e como consequência disso o seu padrão gestual sofreria alterações. Este não parece ser o caso, pois quando estes músicos acompanham outros músicos eles tendem a deslocar o seu padrão gestual em direção ao padrão gestual do outro. Aparentemente este efeito é causado pelas pequenas adequações nos perfis gestuais dos músicos enquanto estes são submetidos a diferentes situações interpretativas (ao acompanhar diferentes *primos*). Assim como demonstrado pelos resultados da discriminação entre *primos*, estas variações nos perfis gestuais não são grandes o suficiente para permitir uma identificação dos *primos* clarinetistas acompanhados. Porém elas parecem ser suficientes para indicar a tendência de se aproximar do outro.

Acreditamos que estas variações estão relacionadas às manipulações dos parâmetros acústicos indicada por cada *primo*, já que durante as gravações não houve interação entre os clarinetistas que não fosse exclusivamente através do sinal de áudio tocado através de fone de ouvido. Quando os *secondos* clarinetistas seguiam os *primos*, eles tentavam adequar sua performance às intenções do outro. Consequentemente, é possível que as alterações observadas nos padrões gestuais sejam um reflexo desta tentativa de adequação. Logo, estes resultados são favoráveis à hipótese de acoplamento entre gesto e música. Diversos estudos apontam evidências desta relação, por exemplo, experimentos realizados por Caramiaux, Bevilacqua e Schnell (2010) demonstraram que parâmetros extraídos dos gestos dos ouvintes (posição, módulo da velocidade, acelerações tangencial e normal, curvatura, raio e torção) foram correlacionados com os parâmetros acústicos (*Loud-*

ness e Sharpness) extraídos dos sons escutados. Em um outro estudo realizado por Dahl e Friberg (2007) os participantes assistiram e classificaram vídeos silenciosos de percussionistas tocando sob diferentes intenções emocionais (feliz, triste, irritado, e com medo). Os resultados demonstraram que os gestos dos percussionistas influenciaram a percepção das intenções expressivas da audiência.

A comunicação humana não se limita ao uso de apenas uma das vias de informação sensorial existentes. Assim como na comunicação verbal, a música faz uso de diversos artifícios não diretamente associados à produção sonora, como por exemplo a movimentação do corpo para a manutenção do andamento desejado ou a informação visual nas expressões faciais e gestuais de um regente. Acreditamos que a recorrência dos padrões de movimentação observada entre as execuções de um mesmo músico seja um forte indicativo da existência de informação musical codificada no gesto.

A investigação das relações multi-modais presentes na performance musical pode no futuro ajudar a clarificar o processo de construção de uma interpretação. É evidente que a informação gestual interfere na interação entre os músicos de um conjunto. Um estudo recente realizado por Goebel e Palmer (2009) demonstrou que ao tocar em conjunto os músicos com realimentação auditiva reduzida se moviam mais do que os outros, e isto teve consequências para a sincronização do grupo. Os nossos resultados demonstram indícios da importância desta relação já que os músicos testados tendem a se aproximar do padrão de movimentação dos outros. Uma maior ou menor adequação ao padrão gestual do outro poderia ter relação com a capacidade de se “ajustar” musicalmente. Por exemplo, músicos com padrões de movimentação similares poderiam ter uma maior facilidade em sincronizar determinados parâmetros acústicos, com a informação gestual atuando como um catalisador da comunicação entre eles. Uma investigação destas relações poderia lançar luz sobre questões como, por que alguns músicos se sentem confortáveis em tocar com determinado músico mas não com outro, e se a informação gestual contribui para isto, mesmo que implicitamente. Da mesma forma que músicos de uma orquestra se “acostumam” com o gestual de regente, é possível que os músicos aprendam a “ler” a movimentação dos colegas. Para o regente este aprendizado por parte dos músicos é essencial para reduzir o tempo necessário para a construção de uma performance, e consequentemente o desgaste do grupo.

5 *Conclusão e perspectivas futuras*

Um dos objetivos deste estudo foi a criação de um *framework* de análise *multi-modal*, capaz de acessar as relações entre gesto e música na performance musical. Em parte, este objetivo foi atingido. Os procedimentos utilizados se mostraram capazes de identificar uma “assinatura” gestual dos intérpretes. Mais ainda, foi demonstrado que estas “assinaturas” são afetadas quando os parâmetros musicais são alterados, por exemplo ao seguir diferentes músicos. Porém a interação entre gesto e música aponta para um relacionamento complexo, dependente de inúmeras variáveis. Este estudo, dentre outros, expõe a necessidade de diversificar a abordagem do problema, talvez focando-se em partes específicas do todo, como por exemplo a segmentação e a parametrização dos dados gestuais. A segmentação de uma performance musical a partir de blocos gestuais poderia demonstrar uma relação destes blocos com estruturas musicais (frases, motivos, etc). A parametrização poderia indicar uma conexão sutil entre parâmetros musicais e acústicos. A hipótese de acoplamento entre gesto e música possui ainda grande potencial na criação de novas interfaces musicais, síntese a partir de padrões cinemáticos e sistemas de reconhecimento de intérpretes a partir de informação gestual. Uma outra aplicação interessante são sistemas capazes de acompanhar uma partitura (*score following*). Estes sistemas poderiam ser ampliados com a utilização de informação gestual de forma a criar um “acompanhador” ou correpetidor virtual, capaz de aprender como o músico organiza a sua performance.

A identificação automática de músicos a partir de seu estilo de performance é uma tarefa que possui grande potencial para a manipulação e gerenciamento de base de dados musicais. No futuro, talvez sejamos capazes de encontrar automaticamente artistas que possuem conceitos estilísticos similares ou mesmo que estudaram com os mesmos professores utilizando apenas informações extraídas do sinal de áudio das suas performances. Por exemplo, um apreciador de perfor-

mances historicamente informadas poderia filtrar suas buscas focando-se na sonoridade específica de um trompete natural, ou no estilo de resolução de trinados sugerido em um tratado de época.

Os resultados encontrados neste estudo são preliminares. Somente a reprodução sistemática dos experimentos e dos procedimentos realizados poderá no futuro validar as suposições feitas. Planejamos em breve expandir este estudo utilizando diferentes excertos musicais. Esperamos que estes resultados associados aos de outros estudos contribuam positivamente para o avanço na compreensão dos mecanismos de sincronização sensório-motor humano.

Referências Bibliográficas

- BELLO, J. et al. A tutorial on onset detection in music signals. *Speech and Audio Processing, IEEE Transactions on*, v. 13, n. 5, p. 1035–1047, 2005.
- BRESIN, R. Articulation rules for automatic music performance. In: *Proceedings of the International Computer Music Conference*. [S.l.: s.n.], 2001. p. 294–297.
- CADOZ, C.; WANDERLEY, M. et al. Gesture-music. *Trends in Gestural Control of Music*, Paris: Ircam, p. 71–94, 2000.
- CARAMIAUX, B.; BEVILACQUA, F.; SCHNELL, N. Towards a gesture-sound cross-modal analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 158–170, 2010. Disponível em: <http://dx.doi.org/10.1007/978-3-642-12553-9_14>.
- CLYNES, M. Microstructural musical linguistics: composers' pulses are liked most by the best musicians. *Cognition*, Elsevier, v. 55, n. 3, p. 269–310, 1995.
- DAHL, S.; FRIBERG, A. Visual perception of expressiveness in musicians' body movements. *Music Perception*, JSTOR, v. 24, n. 5, p. 433–454, 2007.
- DELALANDE, F. La gestique de gould. *Glenn Gould Pluriel*, p. 85–111, 1988.
- DUDA, R.; HART, P. *Pattern classification and scene analysis*. second. [S.l.]: Wiley, 1996.
- FASTL, H.; ZWICKER, E. *Psychoacoustics: facts and models*. [S.l.]: Springer-Verlag New York Inc, 2007.
- FISHER, R. The statistical utilization of multiple measurements. *Annals of Human Genetics*, Wiley Online Library, v. 8, n. 4, p. 376–386, 1938.
- GABRIELSSON, A. Music performance research at the millennium. *Psychology of music*, Sage Publications, v. 31, n. 3, p. 221–272, 2003.
- GOEBL, W.; PALMER, C. Synchronization of timing and motion among performing musicians. *Music Perception*, JSTOR, v. 26, n. 5, p. 427–438, 2009.
- JEPSEN, M.; EWERT, S.; DAU, T. A computational model of human auditory signal processing and perception. *The Journal of the Acoustical Society of America*, v. 124, p. 422, 2008.

JOHNSON, R.; WICHERN, D. *Applied multivariate statistical analysis*. 4. ed. [S.l.]: Prentice hall Upper Saddle River, NJ, 1998.

KELLER, P. Attentional resource allocation in musical ensemble performance. *Psychology of Music*, SAGE Publications, v. 29, n. 1, p. 20–38, 2001.

KELLER, P.; APPEL, M. Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, JSTOR, v. 28, n. 1, p. 27–46, 2010.

KELLER, P.; KNOBLICH, G.; REPP, B. Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and cognition*, Elsevier, v. 16, n. 1, p. 102–111, 2007.

KIM, H.; MOREAU, N.; SIKORA, T. *MPEG-7 audio and beyond*. [S.l.]: Wiley Online Library, 2005.

KNOBLICH, G.; FLACH, R. Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, SAGE Publications, v. 12, n. 6, p. 467–472, 2001.

KNOBLICH, G. et al. Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *The Quarterly Journal of Experimental Psychology: Section A*, Taylor & Francis, v. 55, n. 3, p. 1027–1046, 2002.

LOUREIRO, M. A. et al. Padrões de sincronização temporal em duos de clarinetas: influencia do acompanhante e da estrutura musical. In: *Anais do XVIII Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música*. João Pessoa, Brasil: Universidade Federal do Paraíba, 2012. p. 581–585. Disponível em: <http://www.anppom.com.br/anais/anaiscongresso_anppom_2012-/VIII_MusicaeTecnologiaeSonologia.pdf>.

LOUREIRO, M. A. et al. Content analysis of note transitions in music performance. In: GOUYON, F.; BARBOSA, A.; SERRA, X. (Ed.). *Proceedings of the 6th Sound and Music Computing Conference Sound and Music Computing Conference*. Porto: INESC, 2009. p. 355–359. Disponível em: <<http://smc2009.smcnetwork.org-/programme/pdfs/291.pdf>>.

LU, X.; WANG, Y.; JAIN, A. Combining classifiers for face recognition. In: IEEE. *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*. [S.l.], 2003. v. 3, p. III–13.

MAESTRE, E.; GÓMEZ, E. Automatic characterization of dynamics and articulation of expressive monophonic recordings. In: CITESEER. *Proceedings of the 118th Audio Engineering Society Convention*. [S.l.], 2005.

- MAESTRE, E. et al. Using concatenative synthesis for expressive performance in jazz saxophone. In: CITESEER. *Proceedings of International Computer Music Conference*. [S.l.], 2006.
- MARCIALIS, G.; ROLI, F. Fusion of lda and pca for face verification. *Biometric Authentication*, Springer, p. 30–37, 2006.
- MCADAMS, S. Perspectives on the contribution of timbre to musical structure. *Computer music journal*, MIT Press, v. 23, n. 3, p. 85–102, 1999.
- MION, L.; POLI, G. D. Score-independent audio features for description of music expression. *Audio, Speech, and Language Processing, IEEE Transactions on*, IEEE, v. 16, n. 2, p. 458–466, 2008.
- MOLINA-SOLANA, M.; ARCOS, J.; GOMEZ, E. Using expressive trends for identifying violin performers. In: *Proceedings of ninth international conference on music information retrieval (ISMIR2008)*. [S.l.: s.n.], 2008. p. 495–500.
- MOORE, B. C. J. *An Introduction to the Psychology of Hearing*. 5th. ed. [S.l.]: Academic Press, 2003.
- MOTA, D. A.; CAMPOLINA, T.; LOUREIRO, M. A. Expan: a tool for musical expressiveness analysis. In: *Proceedings of the 2nd International Conference of Systematic Musicology (Sys Mus2009)*. Ghent, Belgium: Ugent, 2009. p. 24–27. Disponível em: <http://www.ipem.ugent.be/sysmus09/files%20-%20Sysmus09_webproceedings.pdf>.
- NETTHEIM, N.; BECKING, G. How musical rhythm reveals human attitudes: Gustav becking's theory. *International Review of the Aesthetics and Sociology of Music*, JSTOR, p. 101–122, 1996.
- PALMER, C. et al. Synchronization of motion and timing in clarinet performance. In: *International Symposium on Performance Science*. [S.l.: s.n.], 2009. p. 1–6.
- PARNCUTT, R. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, JSTOR, p. 409–464, 1994.
- PEETERS, G. A large set of audio features for sound description (similarity and classification) in the cuidado project. IRCAM, 2004.
- PEETERS, G.; RODET, X. Automatically selecting signal descriptors for sound classification. IRCAM, 2002.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2008. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.
- RAICHEL, D. *The science and applications of acoustics*. [S.l.]: Springer, 2006.

- REPP, B. Musical motion: Some historical and contemporary perspectives. In: A. Friberg, J. Iwarsson, E. Jansson, and J. Sundberg (Éd.), *Proceedings of the Stockholm Music Acoustics Conference*. [S.l.: s.n.], 1993. p. 128–135.
- REPP, B. Expressive timing in schumann's "träumerei:" an analysis of performances by graduate student pianists. *The Journal of the Acoustical Society of America*, v. 98, p. 2413, 1995.
- STONE, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 111–147, 1974.
- TODD, M.; NEIL, P. The dynamics of dynamics: A model of musical expression. *The Journal of the Acoustical Society of America*, v. 91, n. 6, p. 3540–3550, 1992.
- TODD, N. A computational model of rubato. *Contemporary Music Review*, Taylor & Francis, v. 3, n. 1, p. 69–88, 1989.
- TODD, N. Towards a cognitive theory of expression: The performance and perception of rubato. *Contemporary Music Review*, Taylor & Francis, v. 4, n. 1, p. 405–416, 1989.
- TODD, N. Motion in music: A neurobiological perspective. *Music Perception*, JSTOR, p. 115–126, 1999.
- VENABLES, W. N.; RIPLEY, B. D. *Modern Applied Statistics with S*. Fourth. New York: Springer, 2002. ISBN 0-387-95457-0. Disponível em: <<http://www.stats.ox.ac.uk/pub/MASS4>>.
- WANDERLEY, M. Quantitative analysis of non-obvious performer gestures. *Gesture and sign language in human-computer interaction*, Springer, p. 241–253, 2002.
- WANDERLEY, M. et al. The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, Taylor & Francis, v. 34, n. 1, p. 97–113, 2005.
- WANDERLEY, M. M.; DEPALLE, P. Gesturally-controlled digital audio effects. In: CITESEER. *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*. [S.l.], 2001.
- WIDMER, G.; GOEBL, W. Computational models of expressive music performance: The state of the art. *Journal of New Music Research*, Taylor & Francis, v. 33, n. 3, p. 203–216, 2004.
- WOLFE, D.; HOLLANDER, M. Nonparametric statistical methods. *Nonparametric statistical methods*, John Wiley New York, 1973.