

Universidade Federal do Rio Grande do Sul  
Ciências Biológicas

Trabalho de conclusão de curso em formato de artigo nos moldes da Revista  
Brasileira de Ornitologia.

**Quem canta igual? Análise quantitativa de semelhança entre vocalizações  
de aves com base em testes de identificação por peritos**

Aluno: Bento Collares Gonçalves  
Orientador: Gonçalo Ferraz

## ABSTRACT

This study aims to uncover patterns of species identification error in aural surveys of Central Amazon birds. Some amount of observer error is unavoidable and some species have more similar vocalizations than others, therefore, those species are especially prone to be confused. To quantify identification errors we developed an on-line quiz, based on vocalizations of an undisclosed set of 41 antbird (*Thamnophilidae*) and woodcreeper (*Dendrocolaptinae*) species and invited experts to answer the on-line quiz. The quiz results, with 820 answers from 20 participants, were compared to the results of a binomial experiment with a probability of success of 0.5; with this approach we examined whether participants got species right or wrong more often than expected at random and whether species were mistakenly identified more often than expected at random – with the given success probability. We also compiled the quiz answers in a triangular matrix showing species ranked by phylogenetic order on both axes. Thus, correct answers appeared on the main diagonal, while any value in the subdiagonal shows a confusion between the species in the  $y$  axis and in the  $x$  axis. From the triangular matrix we can ask whether closely related species were mistaken for each other more often than distant species. To test this hypothesis, we used a Monte Carlo test to compare the mean distance between the mistaken species in the observed matrix with the mean distance of mistaken species in 1000 randomized matrices. The 20 participants who took the quiz showed substantial variation in their ability to identify species correctly. Four observers who got species wrong more often than expected at random were excluded from subsequent analysis. When drawing comparisons across species, we found that 14 were correctly identified significantly more often than expected at random while only one was misidentified significantly more often than expected at random. Finally, the observed mean distance between confused species was smaller than all of the mean distances from the randomized matrices on the Monte Carlo test, indicating that confusions are more frequent between closely related species.

## RESUMO

Este estudo visa identificar padrões de erro de peritos na identificação de espécies em amostragens bioacústicas de aves da Amazônia Central. Como existe grande semelhança entre algumas das vocalizações e qualquer tipo de amostragem está sujeita a erros, a identificação dos erros mais frequentes ajuda a criar modelos mais robustos de presença/ausência de espécies. Para quantificar os erros de identificação nós desenvolvemos um questionário on-line, baseado em um grupo de 41 espécies de tamnofilídeos e dendrocolaptíneos, e enviamos este questionário a um conjunto de peritos. Os resultados do teste (com 820 respostas de 20 participantes) foram comparados com os resultados de um experimento binomial com probabilidade de sucesso igual a 0,5. Esta comparação permitiu averiguar se cada participante errou ou acertou as perguntas com uma frequência diferente do esperado ao acaso (50% de acertos). Da mesma forma, também averiguamos se as espécies são identificadas corretamente ou incorretamente com mais frequência do que o esperado ao acaso. Também mapeamos as confusões entre espécies em uma matriz triangular com a lista de espécies por ordem filogenética nos eixos  $x$  e  $y$ . Nesta configuração, as respostas corretas ficam na diagonal enquanto as confusões entre espécies aparecem na sub-diagonal. Com base na matriz triangular podemos perguntar se as espécies mais aparentadas são confundidas com maior frequência que espécies distantes. Para testar esta hipótese, usamos um teste de Monte Carlo que compara a média das distâncias entre as espécies confundidas na matriz triangular com a distribuição de médias das distâncias entre espécies confundidas em 1000 matrizes randomizadas. A análise do padrão de respostas dos 20 participantes nos permitiu remover do estudo 4 participantes que responderam errado significativamente mais vezes do que o esperado ao acaso. Usando as respostas dos demais participantes e analisando o padrão de erros e acertos para cada espécie, observamos que 14 espécies foram corretamente identificadas significativamente mais vezes do que o esperado ao acaso, enquanto que só uma deu origem a significativamente mais erros do que o esperado ao acaso. Por fim, a média observada de distâncias entre espécies confundidas foi sempre inferior à média de distância obtida a partir das matrizes randomizadas no teste de Monte Carlo, indicando que as confusões são mais frequentes entre espécies aparentadas.

## INTRODUÇÃO

Vivemos em um período de acelerada extinção de espécies, possivelmente a caminho da sexta grande extinção em massa (Wake and Vredenburg 2008). Em face a esta crise, é preciso que se tenha um bom acompanhamento de quais espécies estão presentes em cada local e o estado de suas populações ao longo dos anos. Este acompanhamento da presença e distribuição das espécies pode se dar através de diversas técnicas de amostragem. A amostragem bioacústica (i.e. por gravação de sons) vem se tornando um eficiente meio de obtenção de informação sobre a presença/ausência de espécies com vocalizações conspícuas como anuros, aves e alguns grupos de insetos. Quando comparada com a observação visual, a amostragem bioacústica tem as vantagens de a) reduzir o esforço necessário para a obtenção de dados; b) aumentar a probabilidade de detecção de espécies em áreas onde a visibilidade é baixa; c) permitir amostragem simultânea em todas as direções; d) permitir fácil armazenamento e revisão. Estas vantagens tornaram possível o monitoramento de longa duração em áreas onde outros métodos, como a observação em transectos lineares, são impraticáveis (Dawson and Efford 2009).

Apesar das vantagens citadas acima, a amostragem bioacústica, como toda a amostragem de campo, pode produzir erros, os quais devem ser devidamente considerados nas análises. Os erros de amostragem de presença/ausência de espécies podem se agrupar em dois tipos principais: falsos positivos e falsos negativos. Os falsos negativos são provenientes de falhas de detecção e resultam na afirmação de ausência de uma espécie quando, em realidade, a espécie está presente. O problema acentua-se com espécies raras e/ou incospícuas e tem sido eficazmente tratado por meio de estimativas de ocorrência de espécies que levam em consideração as falhas de detecção (McKenzie et al. 2002). Os falsos positivos, por outro lado, são provenientes de casos onde houve detecção mas a espécie foi identificada erroneamente. Este tipo de erro recebeu menos visibilidade até o presente momento do que os erros de falsos-negativos (Miller et al. 2013). Não obstante, a relevância dos falsos positivos é evidente, pois eles podem causar equívocos em inferências de ocupação, taxas de colonização e taxas extinção mesmo quando presentes em proporções muito pequenas do total de detecções (McClintock et al. 2010; Miller et al. 2011). Estimativas de ocupação que levam em conta falsos-positivos, em casos onde a ocorrência de erros de identificação é presumível, produzem resultados substancialmente diferentes, e com maior suporte estatístico, em comparação com modelos que desconsideram a existência de falsos positivos (Miller et al.

2011). As soluções analíticas para estimar e corrigir os erros de falso positivo ainda se encontram em amadurecimento.

Talvez a causa mais importante dos erros de amostragem por falso positivo com dados de bioacústica seja a semelhança entre as vocalizações de muitas espécies (McClintock et al 2010). Essa causa, evidentemente, é somada às dificuldades causadas por um conjunto de fatores ambientais, como o ruído de fundo, a distância entre o animal e o gravador e a densidade da vegetação, dificuldades que alteram a qualidade do sinal acústico e consequentemente dificultam sua identificação (Pacifici et al. 2008). Devido a estes e outros fatores como a frequente ausência de confirmação visual, a presença de uma parcela de erros de identificação é inevitável (Farmer et al. 2012).

Adicionalmente, é possível que os erros de identificação sejam mais comuns entre espécies aparentadas. Em uma tentativa de prever se estes erros de identificação são mais comuns entre espécies aparentadas, é importante saber se as vocalizações das espécies em questão são inatas ou aprendidas com os pais. Isso porque vocalizações inatas possuem maior herdabilidade e, portanto, são características mais informativas para a construção de uma filogenia (Miller and Baker 2009). Dentre todos grupos de aves, o aprendizado de vocalizações está restrito aos papagaios (Psittaciformes), beija-flores (Apodiformes: Trochilidae) e pássaros oscinos (Passeriformes - Passeri)(Jarvis 2006). Para o grupo de aves abordado nesse estudo, (Passeriformes: Tyranni: Furnariides), acredita-se que as vocalizações são inatas (Isler et al. 1998). Entretanto, estas conclusões são baseadas em inferências e observações anedóticas, faltando dados experimentais para confirmá-lo. No único experimento para identificar se a vocalização da ave adulta é inata ou não realizado com uma espécie do grupo de aves abordado neste estudo, filhotes criados em cativeiro, sem contato com a vocalização de adultos conspecíficos, desenvolveram uma vocalização indistinguível da vocalização de seus conspecíficos selvagens (Touchton et al. 2014). Porém, com a descoberta que os papagaios são o grupo irmão dos pássaros (Jets et al. 2012), é plausível que o aprendizado de vocalizações seja a condição ancestral dos pássaros da infra-ordem Furnariides, pondo em dúvida a origem inata de suas vocalizações (Touchton et al. 2014).

Na resolução do problema dos falsos positivos é importante saber não apenas que houve o erro de identificação, mas também quais foram as espécies confundidas. Sabendo com que frequência os erros de identificação ocorrem entre quais espécies, pode-se separar os dados em grupos com diferentes níveis de incerteza (Miller et al. 2013), que podem receber

tratamentos distintos nas posteriores análises. Neste estudo, procuramos identificar, dentro de um conjunto de espécies de aves amazônicas, quais são mais prováveis de se confundir com outras, quando identificadas por um conjunto de peritos em identificar as espécies focais do estudo. A par deste objetivo geral, procuramos responder a três perguntas específicas: a) Quanta variação existe no desempenho de diferentes peritos de identificação; b) Quanta variação existe no desempenho destes peritos para identificar cada uma das espécies estudadas(Quais espécies são fáceis ou difíceis de se identificar); e c) Até que ponto espécies filogeneticamente mais aparentadas são mais fáceis de confundir entre si.

## **METODOLOGIA**

### **Área de estudo**

Este estudo foi baseado em dados obtidos a partir de um questionário on-line. O questionário, por sua vez, foi baseado em uma coleção de gravações de vocalizações de aves que ocorrem no Escudo das Guianas, na área de estudo do Projeto de Dinâmica Biológica de Fragmentos Florestais (PDBFF). Esta área está localizada a 80km ao norte de Manaus, e tem sido alvo de uma grande variedade de estudos de flora e fauna. Por esse motivo temos acesso a informação relativamente completa sobre a avifauna da área (Cohn-Haft et al. 1997) e a coleções de referência que nos permitem identificar vocalizações de aves com bastante segurança (Naka et al. 2008). Embora a área de estudo tenha cerca de 400 espécies, este estudo enfocará um conjunto de espécies menor com o objetivo de construir um questionário relativamente curto e assim conseguir o envolvimento no estudo de um número de peritos suficientemente grande.

### **Construção de uma biblioteca de vocalizações**

A primeira etapa deste trabalho foi montar uma biblioteca de vocalizações de espécies de aves das famílias *Thamnophilidae* e *Dendrocolaptidae* ocorrentes na área do PDBFF. Ambas as famílias pertencem a infra-ordem *Furnariides*, que por sua vez pertence a sub-ordem *Tyranni* dentro dos *Passeriformes*. Escolhemos as famílias *Thamnophilidae* e *Dendrocolaptidae* por duas razões: a) estes grupos são compostos quase exclusivamente por aves do sub-bosque, portanto mais fáceis de escutar e de gravar; e b) as filogenias de ambos os grupos estão relativamente bem resolvidas (Irestedt et al. 2004; Moyle et al. 2009), o que nos permite perguntar se espécies filogeneticamente mais próximas são mais fáceis de confundir do que espécies filogeneticamente distantes entre si.

As gravações usadas neste estudo foram obtidas na base de dados de gravações autônomas do Laboratório de Ecologia de Populações da UFRGS e da coleção online da *Xeno-Canto Foundation*, disponível gratuitamente no endereço [www.xeno-canto.org](http://www.xeno-canto.org). As gravações foram editadas no programa *Adobe Audition 5.5* para que todas tivessem duração, ruído e amplitude de sinal semelhantes; desta forma procuramos assegurar que as variações na identificação fossem determinadas pelas variações nas características da vocalização e não na qualidade da gravação. Entretanto, para melhor simular uma situação real de processamento

de gravações, não tentamos eliminar completamente ruídos e imperfeições. No total nossa coleção contém 82 vocalizações de 41 espécies (13 dendrocolaptíneos e 28 thamnofilídeos), com duas vocalizações para cada espécie.

### **Questionário On-line**

Para avaliar a possibilidade de confusão entre vocalizações das espécies da biblioteca montamos um questionário on-line usando o programa *Wondershare Quiz Creator*. O questionário consiste de 41 perguntas, selecionadas aleatoriamente do total de 82. Cada pergunta consistiu em uma gravação acompanhada de um sonograma contendo o trecho onde a espécie em questão estava mais conspícua, ajudando a destacá-la dos eventuais ruídos de fundo. Para responder a cada pergunta, o participante precisa escutar uma gravação correspondente e preencher um espaço em branco, onde escreve o nome da espécie que acredita ter escutado em destaque. Como as 41 perguntas são escolhidas ao acaso durante o teste, o número de perguntas por espécie e o número de espécies escutadas estão sujeitos a variação. No entanto, a amostragem de perguntas é feita sem reposição, o que limita o teste a um máximo de 41 espécies e impede que qualquer espécie apareça mais do que duas vezes. Para que o questionário fosse feito de uma só vez, evitando consultas, cada participante teve um tempo limite de 30 minutos para terminá-lo. Antes de enviar o endereço web do teste aos participantes, testamos o questionário em diferentes navegadores de internet de uso corrente. Para evitar que erros de digitação de nome científico fossem tomados por erro de identificação, desenvolvemos um código no pacote R (R Core Team. 2013) que compara as repostas dos participantes com uma lista de espécies da área e corrige erros de digitação em casos onde houve 5 ou menos caracteres de diferença entre a resposta do participante e qualquer uma das espécies da lista. As repostas que não resultaram suficientemente similares a nenhuma espécie da lista, foram assinaladas como espécie de fora da biblioteca e posteriormente verificadas manualmente.

### **Participantes no questionário**

A busca de participantes para o questionário de identificação de vocalizações começou com uma circular de e-mail com uma solicitação de colaboração, instruções para completar o questionário online e uma breve explicação sobre o propósito do estudo. Para tal, fizemos uma lista de potenciais participantes, todos experientes em identificar espécies de aves Amazônicas por suas vocalizações e vários deles com experiência de processamento de



gravações de vocalizações de aves no laboratório de Ecologia de Populações da UFRGS. No total reunimos respostas de 20 participantes que inevitavelmente têm aptidões variáveis para identificar as aves da biblioteca. Alguns participantes têm mais experiência com identificação visual que auditiva e outros conhecem bem as vocalizações de aves amazônicas mas não necessariamente as vocalizações das espécies da área de estudo.

### **Análise dos Dados**

O desempenho de cada participante na identificação de espécies foi mensurado de acordo com a proporção de espécies que ele ou ela identificaram corretamente. Para agrupar os desempenhos entre excepcionalmente bons, médios, ou excepcionalmente ruins, perguntamos se o número de respostas corretas poderia resultar de um experimento binomial com probabilidade de acerto igual a 0.5. Supondo que, perante cada pergunta, o participante lança uma moeda não viciada que de um lado tem a resposta certa e de outro tem uma resposta errada, qual seria a probabilidade  $P$  de o participante obter um resultado tão ou mais extremo do que o obtido no teste? ‘Extremo’ significa com muito mais respostas corretas do que erradas ou com muito mais respostas erradas do que corretas, do que o esperado ao acaso. Uma vez calculada esta probabilidade por meio da fórmula da distribuição binomial implementada no pacote R (R Core Team. 2013), aplicamos um teste de hipótese de duas caudas para testar a hipótese nula de que o participante tem igual probabilidade de responder certo ou errado. Sempre que a probabilidade de obter o mesmo número ou mais respostas certas que o observado foi  $\leq 0.025$ , consideramos que o desempenho foi excepcionalmente bom. Em contrapartida, quando a probabilidade de obter o mesmo número ou mais respostas erradas foi  $\leq 0.025$ , consideramos o desempenho excepcionalmente ruim e excluímos as respostas do observador em causa dos passos subsequentes da análise. Tomamos a decisão de excluir respostas de participantes com desempenho excepcionalmente ruim para tentar direcionar nossa análise para confusões de identificação que se devem principalmente a semelhança entre vocalizações e não necessariamente à falta de contato prévio do observador com as vocalizações em questão. Em todos os casos em que  $P > 0.025$  consideramos que o participante teve um desempenho médio.

Após a comparação do desempenho dos participantes realizamos uma comparação entre espécies usando apenas as respostas dos participantes de desempenho médio e excepcionalmente bom. A comparação entre espécies seguiu a mesma metodologia de análise que a comparação entre observadores, com a diferença de que aqui o número de lançamentos

de moeda na distribuição binomial corresponde ao número total de vezes  $N$  que o questionário apresentou um canto da espécie em causa (para qualquer observador). Como as perguntas do questionário são selecionadas aleatoriamente, o valor de  $N$  foi diferente entre espécies. Neste caso, o teste de duas caudas sobre a distribuição binomial, nos permitiu identificar espécies particularmente difíceis e espécies particularmente fáceis de identificar. Um valor de  $P \leq 0.025$  significa que é altamente improvável obter um número de acertos ou um número de erros igual ou superior ao observado num cenário em que a probabilidade de obter a resposta certa é igual a 0.5.

Encontramos três tipos diferentes de respostas erradas no teste: respostas em branco (em que o participante não responde), respostas erradas com uma espécie que não ocorre na área de estudo, e respostas erradas com uma espécie que ocorre na área de estudo. Com base no terceiro tipo de erro fizemos uma listagem de confusões entre espécies. Na nossa análise, uma resposta em que o participante escreve o nome da espécie  $b$  ao escutar o canto da espécie  $a$ , conta como confusão entre  $a$  e  $b$  da mesma forma que uma resposta em que o observador escreve espécie  $a$  ao escutar o canto da espécie  $b$ . Ou seja, consideramos que as confusões são simétricas. As confusões encontradas nos testes de todos os observadores com desempenho médio ou excepcionalmente bom foram compiladas em uma matriz triangular com a mesma lista de espécie em linhas e em colunas e com o número de vezes que uma espécie da linha foi confundida com uma espécie da coluna na sub-diagonal. Na diagonal, listamos o número de vezes que a espécie correspondente foi identificada de forma correta. As espécies foram ordenadas na matriz triangular por ordem taxonômica, de acordo com a listagem em Remsen et al.(2014). Desta forma, duas espécies em posições subsequentes na matriz (com distância igual a 1 na lista) são filogeneticamente mais próximas que duas espécies separadas por uma ou mais posições na lista. Para investigar se o número de confusões é maior entre espécies filogeneticamente próximas do que distantes, fizemos um teste de Monte Carlo onde comparamos a distância média entre espécies confundidas nas observações com a distribuição de distâncias média entre espécies confundidas em um conjunto de 1000 matrizes randomizadas com a mesma probabilidade de obter confusões por célula mas com as posições das confusões distribuídas ao acaso pelo espaço da sub-diagonal da matriz.

## **RESULTADOS**

Somando os resultados de todos os 20 participantes no teste de identificação, obtivemos um total de 820 respostas. Destas, 469 (57%) foram identificações corretas, 179 (22%) respostas em branco, 128 (16%) foram confusões com espécies que ocorrem na área de estudo e as 44 (5%) restantes foram confusões com espécies de fora da área de estudo.

O desempenho dos participantes no teste variou substancialmente: seis (30%) atingiram uma probabilidade de acerto significativamente superior, e quatro (20%) significativamente inferior ao esperado em um experimento binomial com probabilidade de acerto igual a 0.5. Os restantes 10 participantes não acertaram nem erraram mais do que seria esperado se a probabilidade de obter a resposta certa fosse independente entre perguntas e sempre igual a 0.5 (Tabela 1).

Removendo os participantes com probabilidade de acerto significativamente inferior a 0.5, a tabela de resultados por espécie (Tabela 2) mostra como as espécies variam na sua dificuldade de identificação. Das 41 espécies testadas, 13 (32%) tiveram uma probabilidade de identificação correta significativamente superior, e uma (*Schistocichla leucostigma*) significativamente inferior a 0.5. Para as restantes 27 espécies (66%) não encontramos evidência de diferença entre a probabilidade de identificação errada ou correta. Uma das 41 espécies (*Cymbilaimus lineatus*) foi sempre identificada corretamente por todos os observadores.

O teste de Monte Carlo realizado sobre a matriz de confusão (Figura 1) entre espécies vocalizadoras e espécies nomeadas na resposta revela que as confusões foram mais frequentes em espécies filogeneticamente mais próximas do que entre espécies relativamente distantes. Em uma comparação de médias com 1000 matrizes randomizadas, a média de distâncias observada na matriz de confusões (Figura 1) foi sempre inferior a média da matriz randomizada (Figura 2). As quatro espécies confundidas com mais frequência foram os tamnofílidos *Willisornis poecilotus* e *Myrmotherula menetriesii*, com 6 confusões de um total de 38 vezes em que as duas espécies foram ouvidas (~16%) e *Thamnophilus murinus* e *Thamnophilus punctatus* com 6 confusões de um total de 33 vezes em que as duas espécies foram ouvidas (~18%).

## DISCUSSÃO

Este trabalho revelou que existem confusões e erros de identificação de aves muito frequentes mesmo entre os especialistas mais experientes trabalhando em condições próximas do ideal. Mesmo em um teste de identificação online, onde os participantes podem escutar um som bem gravado e observar o sonograma correspondente, existem falhas inevitáveis. Estas falhas precisam ser entendidas para que seja possível, por um lado, diminuir a frequência das falhas, e por outro, modelar a probabilidade daquelas falhas que não podemos mesmo evitar. Observamos variação considerável na probabilidade de errar entre participantes, mas também houve muita variação na probabilidade de cada espécie ser erroneamente identificada. Isto indica que existem observadores com melhor e pior desempenho, mas também existem espécies substancialmente mais difíceis de identificar do que outras. Por fim, constatamos que, embora existam muitas confusões possíveis, a probabilidade de confusão entre espécies é mais alta entre espécies filogeneticamente próximas do que distantes, mesmo quando procuramos confusões dentro de um conjunto relativamente restrito de espécies como o abordado neste estudo.

Devido à inevitável heterogeneidade de experiências entre os participantes, a grande variação de desempenho entre eles era esperada (Tabela 1), sendo já conhecido que a experiência tem uma influencia sobre a capacidade de identificação que em parte depende do quanto a espécie focal é rara ou comum (Farmer et al. 2012). Adicionalmente, como o nosso teste apresentou um som e um sonograma, é possível que nem todos os participantes tenham usado as mesmas pistas para identificar uma espécie vocalizante. Em particular, participantes mais acostumados a identificar espécies no campo devem prestar mais atenção ao som, enquanto observadores mais acostumados a processar gravações com o auxílio de programas que mostram o sonograma à medida que o usuário houve a gravação, devem se guiar não só pelo som mas também pelo aspecto visual do sonograma. Alguns observadores usam ambas as fontes de informação.

A variação na dificuldade de identificar corretamente as diferentes espécies ajuda a identificar quais espécies podem ser mais confiavelmente estudadas com base em gravações e quais certamente requerem mais cuidados de identificação e de modelagem de erros. Do conjunto de espécies estudadas, destacam-se *Cymbilaimus lineatus* por nunca ter sido confundido com nenhuma outra espécie e *Schistocichla leucostigma* por ser a única espécie com probabilidade de identificação correta significativamente inferior a 0.5. Essas duas espécies podem fornecer indícios de quais características são comuns a vocalizações de fácil

distinção e, em contrapartida, quais características são comuns a espécies particularmente difíceis de identificar.

Nossa expectativa inicial era de que em um conjunto de espécies tão pequeno, com parentesco tão próximo, os riscos de confusão entre espécies fossem independentes do parentesco filogenético. No entanto, surpreendentemente, o teste de Monte Carlo revelou que as confusões de identificação são mais frequentes entre espécies aparentadas, corroborando com o cenário em que as vocalizações das aves da infra-ordem Furnariides são inatas (Isler et al. 1998), como no experimento com *Hylophylax naevioides* (Touchton et al. 2014). Entretanto, para quantificar apropriadamente esta relação entre distância filogenética e confusão seria interessante levar em consideração uma medida de distância filogenética mais precisa, baseada em uma árvore conjunta de *Thamnophilidae* e *Dendrocolaptinae*. A distância medida em posições na matriz é uma aproximação grosseira da distância de parentesco filogenético, pois assume que a distância entre duas espécies de um dado gênero é igual a distância entre a última espécie deste gênero e a primeira espécie do gênero seguinte – o que não é necessariamente verdade.

Embora fosse muito interessante alargar a amostragem deste estudo para um conjunto de observadores e de espécies maior, os resultados que obtivemos aqui já permitem identificar focos de confusão recorrentes entre algumas das espécies estudadas, como foi o caso das confusões entre *Willisornis poecilinotus* e *Myrmotherula menetriesii* e entre *Thamnophilus murinus* e *Thamnophilus punctatus* (Figura 1). O conhecimento de quais confusões são mais prováveis é de grande utilidade na estimativa da probabilidade de falsos positivos em dados de detecção/não-deteção de espécies. Por sua vez, estas estimativas, quando incorporadas em modelos de ocupação de sítios ajudam a obter previsões mais robustas para a distribuição de espécies no espaço (Miller et al 2013).

No futuro, seria útil expandir a análise que fizemos aqui com um número maior de espécies, reduzindo o número de erros por confusão com espécies da área mas fora da amostra e obtendo uma informação mais completa sobre o padrão de erros de identificação. Também seria pertinente comparar os resultados obtidos neste estudo com dados provenientes de estudos com algoritmos de identificação bioacústica automatizada, podendo revelar os cenários onde cada método é superior. Claramente, este tipo de trabalho pode ser útil também em estudos de outros grupos de espécies comumente identificados por sons, como anuros e quirópteros.

**TABELA 1.** Resultados do teste de identificação para os 20 participantes indicando o número de respostas deixadas em branco ('Branco'), respostas com espécies de fora da área de estudo ('Fora'), ou respostas com uma espécie errada mas da área de estudo ('Errado'). A coluna 'Certo' mostra o número de respostas corretas em um total de 41 perguntas por participante. '*P*' indica a probabilidade de se obter um número de acertos tão ou mais extremo que o observado, em um experimento binomial com 41 tentativas e uma probabilidade de acerto igual a 0.5.

<b>Observador</b>	<b>Branco</b>	<b>Fora</b>	<b>Errado</b>	<b>Certo</b>	<b><i>P</i></b>
Observador A*	23	0	8	10	0.0007
Observador B	5	3	9	24	0.1744
Observador C**	0	0	1	40	<0.0001
Observador D	11	0	4	26	0.0586
Observador E*	3	22	7	9	0.0002
Observador F	14	4	7	16	0.1055
Observador G*	25	0	9	7	<0.0001
Observador H	16	1	5	20	0.3776
Observador I*	22	1	8	10	0.0007
Observador J**	4	3	5	29	0.0057
Observador K**	4	1	3	33	<0.0001
Observador L**	0	1	2	38	<0.0001
Observador M**	5	0	7	29	0.0057
Observador N	8	0	10	23	0.2663
Observador O**	1	1	4	35	<0.0001
Observador P	11	1	5	24	0.1744
Observador Q	11	3	4	23	0.2663
Observador R	15	0	4	22	0.3776
Observador S	0	0	14	27	0.0298
Observador T	1	4	12	24	0.1744

\* Probabilidade de acerto significativamente inferior a 0.5 em um teste de duas caudas com  $p = 0.05$ .

\*\* Probabilidade de acerto significativamente superior a 0.5 em um teste de duas caudas com  $p = 0.05$ .

**TABELA 2.** Respostas organizadas por espécie, mostrando o número de vezes que a vocalização teve resposta em branco ('Branco'), errada com espécie de fora da área de estudo ('Fora'), ou errada com espécie da área de estudo ('Errado'). A coluna 'Certo' mostra o número de respostas certas e a coluna 'N' o total de vezes que a vocalização da espécie apareceu em todos os testes. 'P' é a probabilidade de se obter um número de acertos tão ou mais extremo que o observado, em um experimento binomial com *N* tentativas e uma probabilidade de acerto igual a 0.5.

<b>Espécie</b>	<b>Branco</b>	<b>Fora</b>	<b>Errado</b>	<b>Certo</b>	<b>N</b>	<b>P</b>
<i>Euchrepomis spodioptila</i>	3	1	1	12	17	0.0717
<i>Cymbilaimus lineatus</i> **	1	0	0	13	14	<0.0001
<i>Frederickena viridis</i>	1	0	4	14	19	0.0318
<i>Thamnophilus murinus</i> **	0	0	0	18	18	<0.0001
<i>Thamnophilus punctatus</i>	0	0	6	9	15	0.0303
<i>Thamnomanes ardesiacus</i>	6	0	0	16	22	0.0262
<i>Thamnomanes caesi</i>	7	0	3	12	22	0.4159
<i>Isleria guttata</i>	3	1	5	2	11	0.0327
<i>Epinecrophylla gutturalis</i>	4	2	1	10	17	0.3145
<i>Myrmotherula brachyura</i>	7	0	2	13	22	0.2617
<i>Myrmotherula axillaris</i>	0	0	6	8	14	0.3953
<i>Myrmotherula longipennis</i>	0	1	5	6	12	0.6128
<i>Myrmotherula menetriesii</i>	4	0	4	8	16	0.5982
<i>Herpsilochmus dorsimaculatus</i>	4	0	2	7	13	0.5000
<i>Hypocnemis cantator</i>	6	0	0	12	18	0.1189
<i>Cercomacra cinerascens</i> **	2	0	0	18	20	0.0002
<i>Cercomacra tyrannina</i> **	1	2	0	18	21	0.0007
<i>Cercomacra laeta</i>	1	0	2	8	11	0.1133
<i>Sclateria naevia</i> **	0	0	1	8	9	0.0195
<i>Pernostola rufifrons</i> **	0	0	2	15	17	0.0012
<i>Schistocichla leucostigma</i> *	3	0	5	1	9	0.0195
<i>Myrmeciza ferruginea</i> **	6	0	0	18	24	0.0113
<i>Myrmeciza atrothorax</i>	6	0	1	8	15	0.5000
<i>Myrmornis torquata</i> **	0	2	1	10	13	0.0461
<i>Pithys albifrons</i>	6	0	4	6	16	0.2272
<i>Gymnopithys rufigula</i> **	2	1	2	17	22	0.0084
<i>Hylophylax naevius</i>	2	0	3	13	18	0.0481
<i>Willisornis poecilinotus</i>	5	1	5	11	22	0.5841
<i>Certhiasomus stictolaemus</i>	2	4	2	4	12	0.1208
<i>Sittasomus griseicapillus</i> **	0	0	3	18	21	0.0007
<i>Deconychura longicauda</i>	4	0	3	7	14	0.6047
<i>Dendrocincla merula</i>	2	0	5	7	14	0.6047
<i>Dendrocincla fuliginosa</i>	3	1	3	8	15	0.5000
<i>Glyphorynchus spirurus</i> **	1	0	0	14	15	<0.0001

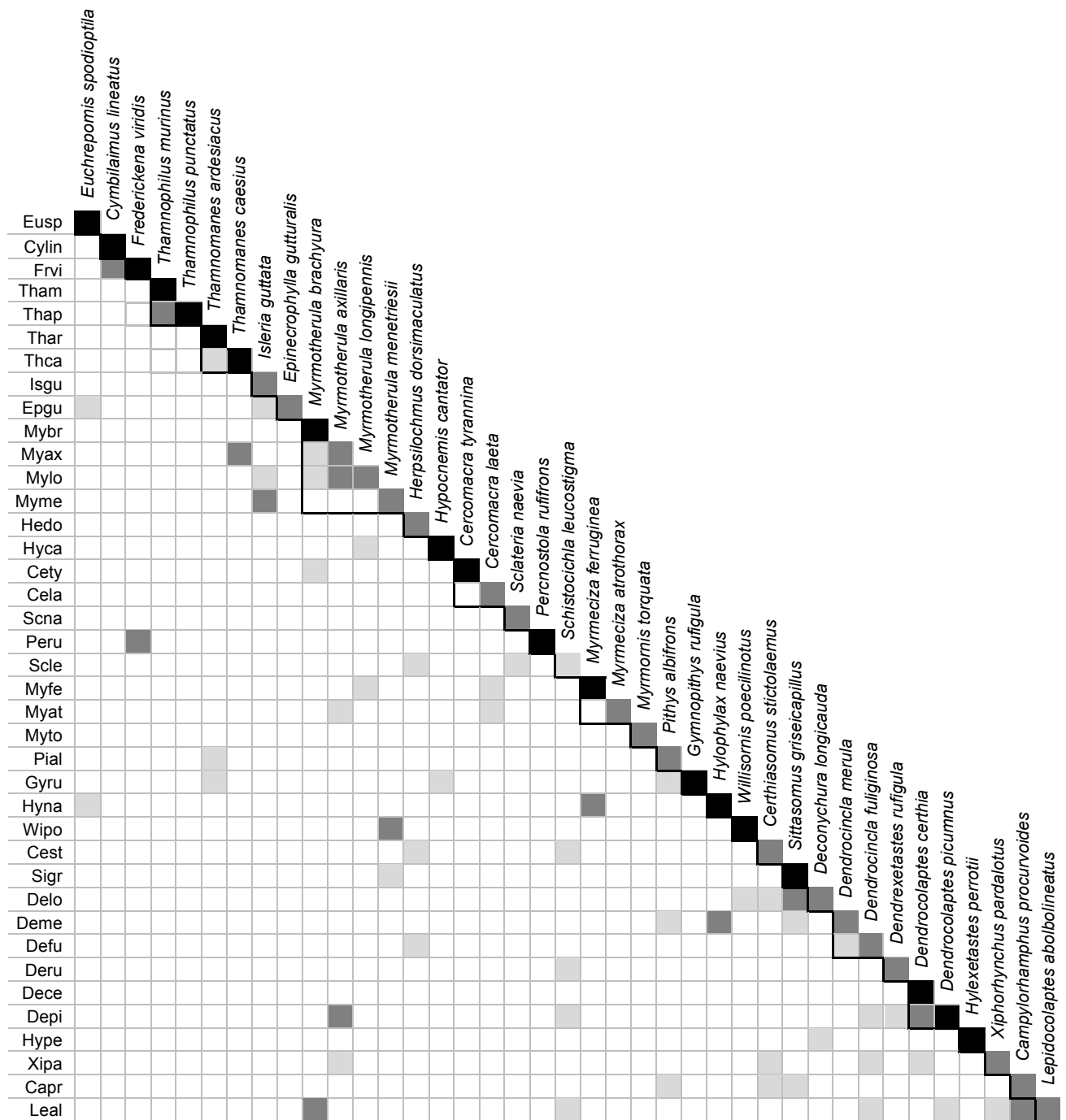
**TABELA 2. (Cont.)**

<b>Espécie</b>	<b>Branco</b>	<b>Fora</b>	<b>Errado</b>	<b>Certo</b>	<b>N</b>	<b>P</b>
<i>Dendrexetastes rufigula</i>	6	2	0	10	18	0.4072
<i>Dendrocolaptes certhia</i> **	1	1	2	13	17	0.0245
<i>Dendrocolaptes picumnus</i>	3	1	3	13	20	0.1316
<i>Hylexetastes perrotii</i> **	0	0	1	11	12	0.0032
<i>Xiphorhynchus pardalotus</i>	2	0	3	10	15	0.1508
<i>Campylorhamphus procurvoides</i>	0	0	1	2	3	0.5000
<i>Lepidocolaptes albolineatus</i>	2	1	5	5	13	0.2905

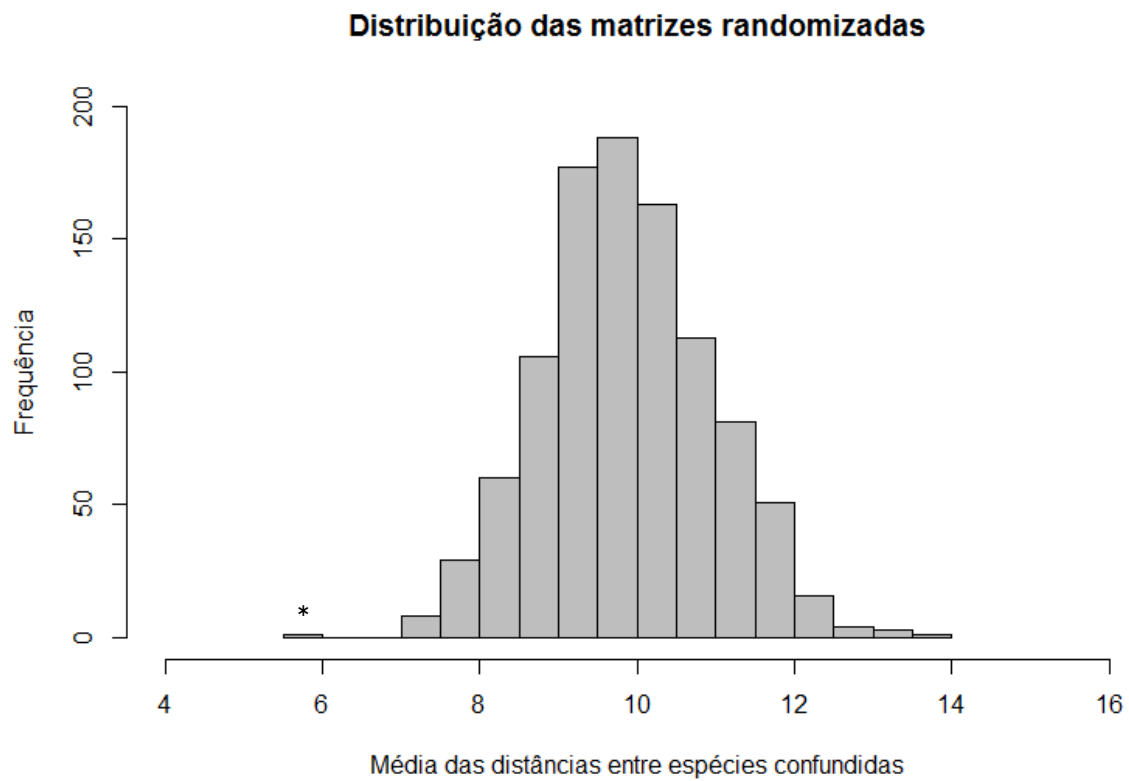
\* Espécie difícil de identificar, com número de erros significativamente diferente do esperado em teste de duas caudas com  $p = 0.05$ .

\*\* Espécie fácil de identificar, com número de acertos significativamente diferente do esperado em teste de duas caudas com  $p = 0.05$ .





**FIGURA 1.** Matriz triangular de respostas ao teste de identificação indicando confusões na sub-diagonal e respostas certas na diagonal. As mesmas espécies aparecem na mesma ordem filogenética em linhas e em colunas, sendo os códigos à esquerda abreviações dos nomes por extenso à direita. A cor de cada célula indica se o número de respostas correspondentes foi 0 (branco), 1 (cinza claro),  $\geq 2$  e  $\leq 10$  (cinza escuro) e  $> 10$  (preto). Confusões entre espécies do mesmo gênero estão destacadas por uma linha preta.



**FIGURA 2.** Histograma das médias de distância entre espécies confundidas nas 1000 matrizes randomizadas na simulação de Monte Carlo. Valores no eixo y indicam a frequência com que as matrizes randomizadas obtiveram média de distância entre espécies confundidas igual ao valor correspondente no eixo x. \* = Média das distâncias entre espécies confundidas observada.

## REFERÊNCIAS CITADAS:

- Cohn-Haft, M., A. Whittaker, and P. C. Stouffer. 1997.** A new look at the "species-poor" Central Amazon: the avifauna north of Manaus, Brazil. *Ornithological Monographs* 48: 205–235.
- Dawson, D. K., and M. G. Efford. 2009.** Bird population density estimated from acoustic signals. *Journal of Applied Ecology* 46: 1201–1209.
- Farmer, R. G., M. L. Leonard, G. Horn, and Andrew. 2012.** Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. *Auk* 129: 76–86.
- Isler, M. L., P. R. Isler, and B. M. Whitney. 1998.** Use of Vocalizations to Establish Species Limits in Antbirds (Passeriformes: Thamnophilidae). *The Auk*:577-590.
- Irestedt, M., J. Fjeldsá, J. A. Nylander and P. G. Ericson. 2004.** Phylogenetic relationships of typical antbirds (*Thamnophilidae*) and test of incongruence based on Bayes factors. *BMC Evolutionary Biology* 4: 23–39.
- Jarvis, E. D. 2006.** Evolution of brain structures for vocal learning in birds: a synopsis. *Acta Zoologica Sinica*:85-89.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012.** The global diversity of birds in space and time. *Nature* 491:444-448.
- McClintock, B. T., L. L. Bailey, K. H. Pollock, and T. R. Simons. 2010a.** Experimental investigation of observation error in anuran call surveys. *Journal of Wildlife Management* 74: 1882–1893.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002.** Estimating Site Occupancy Rates When Detection Probabilities are Less Than One. *Ecology*:2248-2255.
- Miller, E. H. B., Allan J. 2009.** Antiquity of Shorebird Acoustic Displays. *The Auk*:454-459.
- Miller, D. A. W. , J. D. Nichols, B. T. McClintock, and E. H. C. Grant 2011.** Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* 92: 1422–1428.
- Miller D. A. W., J. D. Nichols, J. A. Gude, L. N. Rich, K. M. Podrutzny, J. E. Hines, and M. S. Mitchell. 2013.** Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. *PLoS ONE* 8: e65808.
- Moyle, R. G., R. T. Chesser, R. T. Brumfield, J. G. Tello, D. J. Marchese and J. Cracraft. 2009.** Phylogeny and phylogenetic classification of the antbirds, ovenbirds, woodcreepers, and allies (Aves: Passeriformes: infraorder Furnariides). *Cladistics* 25: 1–20.
- Naka, L. N., P. C. Stouffer, M. Cohn-Haft, C. A. Marantz, A. Whittaker, and R. O. Bierregaard, Jr. 2008.** *Voices of the Brazilian Amazon, Vol. 1. Birds of the terra firme forests north of Manaus: Guianan area of endemism.* Editora INPA, Manaus, Brazil.
- Pacifici, K., T. R. Simons, and K. H. Pollock. 2008.** Effects of vegetation and background noise on the detection process in auditory avian point-count surveys. *Auk* 125:600–607.

- R Development Core Team. 2013.** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 3.0.2. <http://www.R-project.org/>.
- Remsen, J. V., Jr., C. D. Cadena, A. Jaramillo, M. Nores, J. F. Pacheco, J. Pérez-Emán, M. B. Robbins, F. G. Stiles, D. F. Stotz, and K. J. Zimmer. 2014.** A classification of the bird species of South America. American Ornithologists' Union. Version [20/05/2014] <http://museum.lsu.edu/~Remsen/SACCBaseline.html>.
- Touchton, J. M., N. Seddon, and J. A. Tobias. 2014.** Captive rearing experiments confirmsong development without learning in a tracheophone suboscine bird. PLoSOne 9:e95746.
- Wake, D. B. and V. T. Vredenburg. 2008.** Are we in the midst of the sixth mass extinction? A view from the world of amphibians. Proceedings of the National Academy of Sciences of the United States of America 105: 11466–73.