

## **DIMENSIONAMENTO DE UM CALL CENTER: SIMULAÇÃO OU TEORIA DE FILAS?**

**MARCO AURÉLIO CARINO BOUZADA** ( [marco.bouzada@estacio.br](mailto:marco.bouzada@estacio.br) , [marcoareliobouzada@yahoo.com.br](mailto:marcoareliobouzada@yahoo.com.br) )  
*UNIVERSIDADE ESTÁCIO DE SÁ / MADE*

### **RESUMO**

Este trabalho procura estabelecer uma dicotomia – opondo métodos analíticos (como a Teoria de Filas) a métodos experimentais (como a Simulação) – ambientada no problema de dimensionamento da capacidade de atendimento de um grande *call center* brasileiro.

A bibliografia acerca da aplicação de tais métodos em centrais de atendimento é revista e a maneira como o problema em questão é tratado atualmente é descrita em detalhes.

Uma metodologia alternativa à forma tradicional de atacar o problema é sugerida e os resultados obtidos são utilizados para justificar a adequabilidade da abordagem experimental a operações complexas, como a de um *call center* moderno.

**Palavras-chave:** Call center, dimensionamento, Simulação, Teoria de Filas.

### **1. Introdução**

*Call centers* consistem em centros operacionais, instalados para utilizar telecomunicação e tecnologias computacionais, para automatizar uma quantidade volumosa de diferentes atividades e serviços telefônicos, tanto recebidos quanto originados pelo centro. Os centros do tipo *inbound*, onde as chamadas são originadas pelos clientes, são caracterizados como um sistema constituído de várias pessoas (atendentes), que recebe ligações telefônicas de outras pessoas, normalmente clientes – ainda que potenciais, que desejam obter alguma informação, contratar algum serviço, comprar algum produto, responder a alguma pesquisa, atualizar dados, registrar ocorrências, ou fazer reclamações, entre outras demandas. (GROSSMAN et al., 2001; HAWKINS et al., 2001)

Dezenas de bilhões de dólares foram gastos em *call centers* na última metade da década de 1990. O crescimento dessa indústria naquela década foi na faixa de 20% ao ano e é esperado que essa taxa se mantenha neste patamar durante o início desse século. Em particular no Brasil, a indústria de *call centers* tem crescido bastante nos últimos anos, fazendo o mercado doméstico já movimentar acentuadas cifras financeiras. Na economia de hoje, os *call centers* não apenas se tornaram os pontos primários de contato entre clientes e empresas, mas também um grande investimento para muitas organizações. (GROSSMAN et al., 2001; WEINBERG; BROWN; STROUD, 2006; BOUZADA, 2006)

No caso particular de vendas e serviços, *call centers* são frequentemente o ponto de contato entre a empresa e os clientes, o que os torna críticos e caros. Empresas que não tomam as providências necessárias para melhorar a eficiência de seus *call centers* rapidamente descobrem que seus erros de planejamento são traduzidos em baixos níveis de serviço, vendas perdidas, aumentos nos custos e clientes frustrados e insatisfeitos. (BATT; DOELLGAST; KWON, 2004)

O custo de pessoal representa aproximadamente 70% do total da indústria, justificando a necessidade de um gerenciamento eficiente e a importância de uma abordagem quantitativa para o dimensionamento da capacidade de atendimento, que consiste no *trade-off* entre esse custo e a determinação do nível ótimo de serviço; ou seja, em ter o número certo de pessoas qualificadas e recursos de suporte no momento correto, para lidar com a carga de trabalho prevista, mantendo os padrões de qualidade e o nível de serviço exigido. Por isso e à medida que os *call centers* têm apresentado operações cada vez mais complexas, tem se tornado ainda mais importante o uso de modelos acurados para o dimensionamento do tamanho da equipe em uma indústria que movimenta tanto volume financeiro. (HALL; ANTON, 1998; ALAM, 2002; BOUZADA, 2006)

Segundo Hall e Anton (1998), os *call centers* podem usar uma ferramenta conhecida como Simulação para testar (e, eventualmente, justificar a implantação) se determinadas mudanças serão capazes de melhorar o sistema, antes de implantá-las. Os melhores *call centers* usam essa ferramenta efetiva e eficientemente para projetar o sistema, gerenciar a operação e se planejar para o futuro, diante de possíveis cenários.

De acordo com Bouzada (2006), isso ocorre porque, entre outros motivos, o dimensionamento da capacidade de atendimento consiste em uma atividade crítica para a obtenção da eficiência e da efetividade da operação. E a ferramenta de simulação costuma se adequar melhor ao dimensionamento de operações mais complexas (como as de um *call center* moderno), já que ela consegue modelar muito bem a realidade, apresentando resultados mais acurados e relativamente precisos. É bem verdade que esses resultados não são tão precisos quanto os resultados teóricos (obtidos por métodos analíticos), mas representam, normalmente, uma boa aproximação para os mesmos. A máxima da simulação diz que “é melhor ter uma solução aproximada para um modelo bem realista do que uma solução exata para um modelo com muitas aproximações”.

O autor ainda argumenta que uma das principais vantagens da simulação em relação aos métodos analíticos consiste justamente na modelagem mais acurada da realidade: os métodos analíticos precisam fazer muitos ajustes na realidade para que esta se encaixe em algum padrão pré-concebido (exemplo: os tempos de atendimento devem seguir uma distribuição exponencial, o que, muitas vezes, está longe de corresponder à realidade) que a teoria analítica seja capaz de tratar, enquanto a simulação é capaz de representar muito bem quase qualquer peculiaridade de uma operação complexa (por exemplo: roteamento das chamadas, diversidade de funções dos atendentes, abandono de chamadas e outras características comuns nos *call centers* e que, dificilmente, poderiam ser contempladas por uma abordagem analítica). Naturalmente, essa vantagem cresce juntamente com a complexidade da operação estudada e praticamente inexistente em operações muito simples.

Pode parecer contraditório dizer que a simulação consiste em uma abordagem menos precisa e mais acurada, quando comparada aos métodos analíticos. Mas, na verdade, é isso mesmo que ocorre: como os resultados são empíricos, eles variam a cada vez que a simulação é executada, gerando um pouco de imprecisão; entretanto, como estes resultados se baseiam em um modelo mais próximo da realidade, eles são realmente mais acurados (mais próximos dos “valores corretos” para aquele problema real). (BOUZADA, 2006)

Segundo Mehrotra e Fama (2003) e Hall e Anton (1998), *call centers* são interessantes objetos de estudo da simulação, pois: (i) lidam com mais de um tipo de chamada, onde cada tipo representa uma fila; (ii) as chamadas recebidas em cada fila chegam aleatoriamente ao longo do tempo; (iii) em alguns casos, agentes fazem ligações proativamente (tipicamente em *telemarketing* ou ações de cobranças) ou como retorno de uma chamada anterior; (iv) a duração de cada chamada é aleatória, assim como a do trabalho que o agente executa depois da chamada (por exemplo: entrada de dados, documentação, pesquisa); (v) o avanço nos

sistemas para rotear as chamadas para os agentes, grupos ou localizações torna a lógica por trás do *call center* cada vez mais sofisticada; (vi) agentes podem ser treinados para responder a um tipo único de chamada, vários tipos de chamadas ou todos os tipos de chamadas, com diferentes prioridades e preferências especificadas nas lógicas de roteamento; e (vii) a grande quantidade de dinheiro investido em *call centers*, tanto na forma de capital como na forma de trabalho, é capaz de justificar o uso dessa ferramenta tão poderosa.

Com os objetivos de esclarecer a questão e de mostrar as limitações de métodos analíticos para lidar com as peculiaridades reais e de difícil tratamento dos *call centers* (caracterizando as lacunas a serem preenchidas pela abordagem da simulação), Bouzada (2006, p. 7-9) apresenta um exemplo hipotético.

Em vista disso, esse trabalho procura descrever o problema de dimensionamento da capacidade de atendimento de uma grande empresa brasileira de *call centers* para, a seguir, atacá-lo a partir de uma metodologia alternativa: a Simulação. O objetivo é usar o caso da empresa em questão como pano de fundo empírico para a discussão teórica acerca da adequação de métodos experimentais (como a Simulação) a operações complexas (como a dos *call centers* modernos), em detrimento de métodos analíticos (como a Teoria de Filas).

## 2. Referencial teórico

### 2.1. Teoria de Filas aplicada a Call Centers

Em um sistema de um *call center*, uma fila acontece quando não existe um agente disponível para atender um cliente, sendo este colocado em espera em uma fila virtual, da qual ele só sai quando um operador é alocado para atendê-lo ou quando ele abandona a chamada. Conforme observado por Brown et al. (2002), no caso dos *call centers*, a fila virtual formada é invisível entre os clientes (que não vêem uns aos outros) e entre eles e os operadores que os servem.

No cenário de *call centers*, Araújo, Araújo e Adissi (2004) dizem que as disciplinas das filas, se bem gerenciadas, são fortes aliadas da área de planejamento e controle da produção dos *call centers*, que tem como meta atingir os resultados esperados com recursos muitas vezes escassos, tornando esta área cada vez mais importante nestas empresas. As disciplinas das filas, quando bem administradas, podem trazer reduções significativas nos tempos de espera dos clientes que por elas passam.

Algumas características das centrais de atendimento tornam difícil a aplicação de fórmulas analíticas da teoria de filas para a sua modelagem, incluindo: distribuições genéricas para o tempo de atendimento, taxas de chegada variando com o tempo, sobrecargas temporárias e abandonos. O modelo apresentado por Chassioti e Worthington (2004) consiste em uma abordagem prática capaz de incorporar a maioria dessas características.

Segundo Bapat e Pruitte Jr. (1998), as premissas adotadas nas análises baseadas nos modelos analíticos de teoria das filas são extremamente limitadas quando trabalhadas no contexto atual dos *call centers* porque: (i) as chamadas recebidas são todas do mesmo tipo; (ii) a partir do momento em que uma chamada entra em uma fila, ela nunca a abandona – o que acaba, comumente, superestimando a necessidade de mão-de-obra e aumentando os custos de pessoal da empresa; (iii) os agentes atendem às chamadas fundamentados na disciplina onde o “primeiro que entra é o primeiro que sai” (PEPS; ou FIFO, em inglês); e (iv) cada operador atende todas as chamadas da mesma maneira.

Essas premissas raramente são válidas no ambiente em que as empresas de *call center* estão inseridas, já que, de acordo com os autores – dependendo da tolerância individual para esperar o seu atendimento – um cliente pode abandonar a ligação, caso seja colocado em uma fila. Além disso, os operadores normalmente diferem em relação às habilidades e também ao tempo de atendimento de uma chamada. Adicionalmente, as necessidades dos clientes são

muito diferenças e, algumas vezes, uma priorização – de forma que um serviço melhor possa ser prestado – se faz necessária. Mas muitas empresas continuam apoiando as normalmente complexas tomadas de decisão acerca da alocação de recursos através dos modelos analíticos de teoria das filas, motivadas pelas relativas facilidade e rapidez da abordagem.

Muitas centrais de atendimento apresentam uma distribuição genérica (lognormal, por exemplo) para os seus tempos de atendimento, e não necessariamente uma distribuição exponencial negativa (BROWN et al., 2002). O formato exponencial é usado em grande parte da literatura da teoria de filas – tanto para tempo entre chegadas de clientes como para tempo de atendimento – pelo fato de existirem – para as situações onde esses tempos são considerados como variáveis que seguem uma distribuição exponencial – soluções analíticas para o estado estacionário do sistema.

Mas no caso das centrais de atendimento existentes no mundo real, pelo menos a taxa de chegada de clientes varia com o tempo; essa variação é motivada por propagandas, horário de trabalho etc.. A fadiga dos atendentes pode gerar uma variação também no tempo de atendimento à medida que o tempo passa, mas esta é insignificante quando comparada à variação das taxas de chegada. As soluções encontradas na literatura para lidar com taxas de chegada variando em função do tempo não são muito utilizáveis porque envolvem funções de Bessel, de difícil aplicação na prática. (CHASSIOTI; WORTHINGTON, 2004)

## 2.2. Simulação em *Call Centers*

Chokshi (1999), Klungle e Maluchnik (1997), Hall e Anton (1998), Mehrotra e Fama (2003), Avramidis e L'Ecuyer (2005), Klungle (1999) e Bapat e Pruitt Jr. (1998) destacam alguns fatores recentes que contribuíram para o aumento da demanda pelo uso da ferramenta de simulação no setor de *call centers*: (i) a crescente importância dos *call centers* para boa parte das corporações (devido ao rápido crescimento da era da informação e das comunicações e do aparato tecnológico), aumentando a necessidade de utilização de metodologias científicas de tomada de decisão e ferramentas para a sua gerência estratégica em substituição ao uso da intuição; (ii) a complexidade crescente do tráfico de chamadas unido a regras cada vez mais presentes de roteamento baseado em habilidades; (iii) a incerteza cada vez mais predominante nos problemas de decisão normalmente encontrados no gerenciamento operacional das centrais; (iv) rápidas mudanças nas operações e incremento nas atividades de reengenharia resultantes do aumento de fusões e aquisições, volatilidade do negócio, opções de terceirização e utilização de diferentes canais para atingir o consumidor (telefone, *e-mail*, *chat*); e (v) a disponibilidade e o preço acessível dos computadores, aliados a uma gama de aplicações de simulação em *call center* disponíveis no mercado, cada vez mais intuitivas e fáceis de serem aprendidas e usadas.

A simulação, de acordo com Mehrotra (1997), modela explicitamente a interação entre chamadas, rotas e agentes, assim como a aleatoriedade das chegadas de chamadas individuais e dos tempos de atendimento. Através do uso da simulação, gerentes e analistas traduzem dados brutos de *call centers* (previsão de chamadas, distribuições de tempos de atendimento, horários e habilidades de agentes, vetores de roteamento de chamadas etc.) em manuseável informação sobre níveis de serviço, abandono de clientes, utilização de agentes, custos e outras importantes medidas de performance em um *call center*.

Segundo Chokshi (1999) e Klungle e Maluchnik (1997), a utilização da simulação para auxiliar a tomada de decisões no gerenciamento de *call centers* permite os seguintes benefícios: (i) visualizar processos futuros e ser usada como uma ferramenta de comunicação; (ii) validar premissas de processos antes de sua implementação; (iii) analisar detalhadamente o impacto de mudanças (análise de cenários); (iv) prever necessidades agregadas de recursos e

programar a força de trabalho; (v) mensurar indicadores de performance; e (vi) estimar impactos nos custos e economias.

Uma das utilidades da simulação em um *call center*, de acordo com Hall e Anton (1998), é a de avaliação, quando se pode verificar “onde o *call center* está”. A pergunta-chave é “quão eficiente e efetivo é a operação atualmente?”. O objetivo da avaliação é estabelecer uma linha de partida (e de referência) para a mudança.

Segundo Mehrotra, Profozich e Bapat (1997), Yonamine (2006), Gulati e Malcolm (2001), Bapat e Pruitt Jr. (1998) e PARAGON (2005), um modelo de simulação pode (e tem sido cada vez mais usado para tal) – além de normalmente permitir gráficos e animações – contemplar alguns aspectos críticos dos *call centers* receptivos modernos de todos os tamanhos e tipos, entre outros: (i) nível de serviço específico; (ii) flexibilidade na distribuição dos tempos entre chegadas e de atendimento; (iii) consolidação de centrais; (iv) roteamento baseado em habilidades; (v) tipos múltiplos de chamadas; (vi) filas simultâneas; (vii) padrões de abandono de chamadas; (viii) retorno de ligações; (ix) *overflow* e transbordo; (x) priorização de filas; (xi) transferência de chamadas e teleconferências; (xii) preferências, proficiência, tempo de aprendizado e esquema de horário dos operadores. Os *outputs* do modelo podem vir na forma de tempo de espera, taxa de abandono de chamadas (ambos com possibilidade de diferenciação por tipos de chamada) e nível de utilização dos operadores (com possibilidade de diferenciação por tipos de operador). E devido à adequação desta abordagem às características reais e complexas dos *call centers*, a simulação pode tornar mais confiável o seu dimensionamento e gerenciamento.

De acordo com Mehrotra, Profozich e Bapat (1997), Steckley, Henderson e Mehrotra (2005), PARAGON (2005), Mehrotra (1997), Klungle e Maluchnik (1997), Pidd (1998) e Tanir e Booth (1999), os métodos tradicionais mais usados no gerenciamento e dimensionamento de um *call center* (estimativas intuitivas, cálculos improvisados, planilhas e modelos teóricos de filas Erlang) estão se tornando significativamente limitados por causa da variabilidade das chegadas de chamadas, das rotas e do tempo gasto nas ligações, das habilidades e prioridades dos operadores, da heterogeneidade das chamadas e da interação entre elas e os troncos, da dinâmica de abandono das chamadas, das tendências recentes (como o roteamento baseado em habilidades, canais eletrônicos e manuseio interativo de chamadas) e, de uma forma geral, da sofisticação e complexidade cada vez mais presentes nos sistemas de *call centers*. Por exemplo, os modelos analíticos comumente assumem que a chegada dos clientes consiste em um processo Poisson quando, na verdade, os dados de *call centers* rejeitam esta premissa frequentemente. Adicionalmente, planilhas e modelos Erlang superestimam a quantidade de atendentes, além de terem pouca precisão para centrais com atendimento diferenciado.

A simulação estende a capacidade das ferramentas analíticas e operacionais e consiste em uma ferramenta superior quando não está disponível nenhum modelo teórico tratável que forneça uma representação razoável do sistema sob estudo e as médias não são suficientes, a acurácia é importante, a operação é detalhada, a demanda varia muito, gargalos e necessidades de mudanças no desenho dos processos precisam ser identificados e/ou uma animação é necessária para a melhor comunicação de uma modificação para a diretoria. As recentes tendências da indústria demandam abordagens mais sofisticadas e a simulação fornece as técnicas necessárias para se adquirir *insights* sobre essas novas tendências e ajuda a dar forma aos seus desenhos atuais e futuros, consistindo no único método de análise apto a modelar um *call center* eficiente e acuradamente, através de uma abordagem muito mais prática, flexível em termos de *inputs* e *outputs*, e capaz de permitir a inclusão de detalhes importantes, refletir muito melhor a realidade (sem grandes necessidades de simplificações, como nos modelos teóricos), viabilizar um melhor e mais profundo entendimento acerca dos



processos da central e gerar resultados muito mais robustos a respeito da performance do *call center*, permitindo sua otimização de forma bem mais confiável. (PARAGON, 2005; RILEY, 2005; MEHROTRA, 1997; KLUNGLE; MALUCHNIK, 1997; TANIR; BOOTH, 1999; SALIBY, 1989; HILLIER; LIEBERMAN, 1995; HERTZ, 1980; MEHROTRA; PROFOZICH; BAPAT, 1997; BAPAT; PRUITTE JR., 1998; CHOKSHI, 1999; KLUNGLE, 1999; WORTHINGTON; WALL, 1999; RAGSDALE, 2001; MEHROTRA; FAMA, 2003)

### **3. O caso**

#### **3.1. A empresa**

A Contax surgiu no final de 2000, como extensão natural dos negócios da Telemar, em um setor da economia que pouco investia em tecnologia e qualificação de atendimento, para ajudar seus clientes na gestão operacional do serviço de atendimento, agregando valor no relacionamento com os consumidores finais. (CONTAX, 2006)

No Brasil, atualmente, é a empresa de maior crescimento nessa indústria, tendo crescido quase 60% em 2005, com um faturamento de R\$ 1.129 milhões. Caracteriza-se como a maior empresa do ramo em número de posições de atendimento, e a segunda maior em termos de faturamento e número de funcionários, dentro do território nacional. (OUTSOURCING, 2005)

A Contax é uma empresa de capital 100% nacional e hoje opera com mais de 22 mil posições de atendimento, quase 50 mil funcionários e mais de 40 clientes, sendo o principal a própria Telemar, que responde por aproximadamente 60% do faturamento. Os principais produtos relacionados a este cliente são: (i) Velox; (ii) 103; (iii) Suporte técnico e reparo de defeitos; (iv) OI; e (v) 102, que recebe ligações dos clientes que buscam auxílio à lista telefônica.

#### **3.2. Processo atual de dimensionamento da capacidade de atendimento**

O dimensionamento consiste no estudo para adequar as estruturas físicas, técnicas e de pessoal do *call center* aos objetivos da operação de atendimento ao cliente e tem início a partir da previsão intradiária de demanda.

O produto 102 foi o escolhido para ilustrar o problema do dimensionamento, visto que sua demanda é a mais previsível, de forma que seja possível mensurar isoladamente a qualidade do processo de dimensionamento (partindo da premissa que o *input* – previsão de demanda – é de boa qualidade).

O nível de serviço para este produto diz respeito ao tempo de espera do cliente final na linha telefônica, desde o momento de entrada da ligação até ela ser atendida. Em outras palavras, é o tempo que o cliente passa na fila, ouvindo a música de espera e aguardando o operador. Mais precisamente, o nível de serviço consiste no percentual de chamadas atendidas – dentre as completadas, apenas – em até 10 segundos (ou melhor, que aguardam no máximo 10 segundos para serem atendidas).

Como apenas as ligações efetivamente atendidas entram no cálculo do nível de serviço, os abandonos não são contabilizados (nem punidos, portanto), para efeitos de nível de serviço. Mas eles são mensurados através de outro indicador (taxa de abandono) e a Contax é multada quando essa taxa ultrapassa os 2%, em algum mês. Em vista dessa possibilidade, evitar o abandono costuma ser visto como prioridade, um pouco em detrimento do nível de serviço, desde que esse se mantenha acima de patamares mínimos. O nível de serviço não envolve exigências formais no contrato (como a taxa de abandono), mas influencia a relação e a dignidade comerciais; ou seja, é interessante não priorizar apenas o abandono, deixando de se preocupar em manter o nível de serviço em valores decentes.

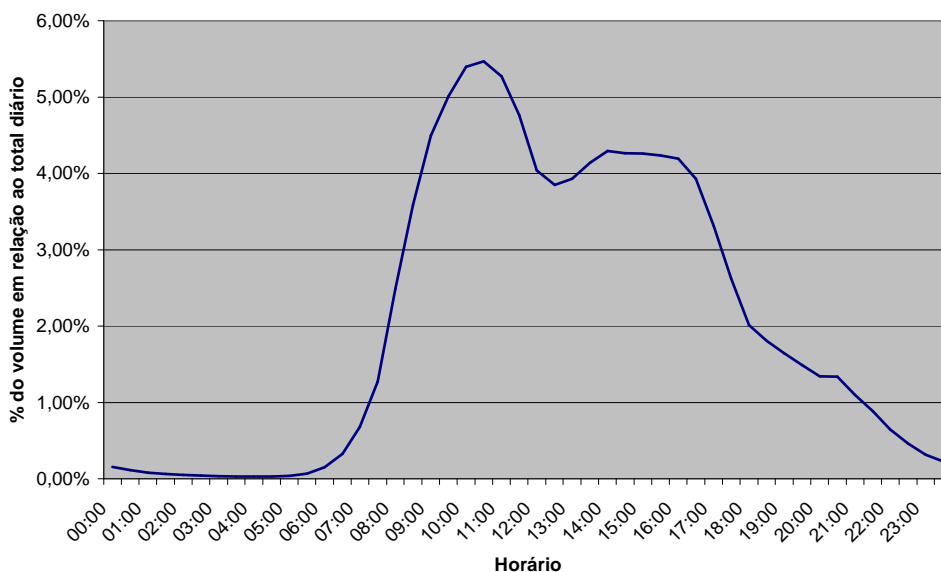
A rotina de dimensionamento inicia com o cálculo das necessidades diárias de atendentes, a partir da previsão da quantidade de ligações, do tempo médio de atendimento e da taxa de ocupação dos atendentes em cada dia. Em seguida, essa necessidade de atendentes (convertida para o padrão de operadores de 6 horas) é comparada com a disponibilidade atual de recursos de atendimento, descontando as perdas referentes a férias e absenteísmo. O resultado dessa comparação é um saldo ou um déficit de mão-de-obra para cada dia do mês planejado. O *output* deste primeiro passo é a quantidade de operadores que precisam ser contratados ou demitidos no mês em questão para os números desejados serem atingidos.

A partir do momento em que a decisão de contratação ou demissão foi tomada e implementada, a equipe de planejamento pode partir para uma análise mais pormenorizada – o dimensionamento intradiário. Este deve ser feito para um dia apenas e esse formato-padrão repetido para os demais dias do período, porque o horário de cada funcionário deve ser o mesmo em todos os dias do mês. É possível, entretanto, conseguir um pouco mais de flexibilidade em relação a isso através do uso de banco de horas.

Em vista disso, um volume de ligações e um tempo médio de atendimento (grandezas necessárias para o dimensionamento) específicos precisam ser escolhidos para serem usados como padrão para o dimensionamento de todos os dias do mês. O dia escolhido para o padrão é, normalmente, o quinto dia de maior movimento. Dessa forma, o dimensionamento vai garantir o nível de serviço desejado para esse dia e todos os dias menos movimentados, mas não para os 4 dias de maior demanda, quando haverá perda de nível de serviço. Isso não representa um problema porque o contrato do produto 102 exige um específico nível de serviço mensal, e não diário.

Em cima do dia escolhido como padrão de dimensionamento para o mês, é aplicada uma curva que reflete o perfil intradiário da demanda, ou seja, que percentual do volume diário acontecerá na primeira meia-hora do dia, na segunda meia-hora, ..., e na última meia-hora do dia. Essa curva é obtida com base no histórico de ligações recebidas a cada intervalo de meia-hora, para cada dia da semana. No caso do produto 102, as curvas de cada dia da semana são muito parecidas (principalmente de segunda à quarta-feira, com pequeno aumento vespertino no volume da quinta e da sexta-feira), sendo as do sábado e do domingo um pouco diferentes. A figura 1 a seguir ilustra a curva histórica da terça-feira.

**Figura 1 – Perfil histórico do comportamento intradiário da demanda, terça-feira**

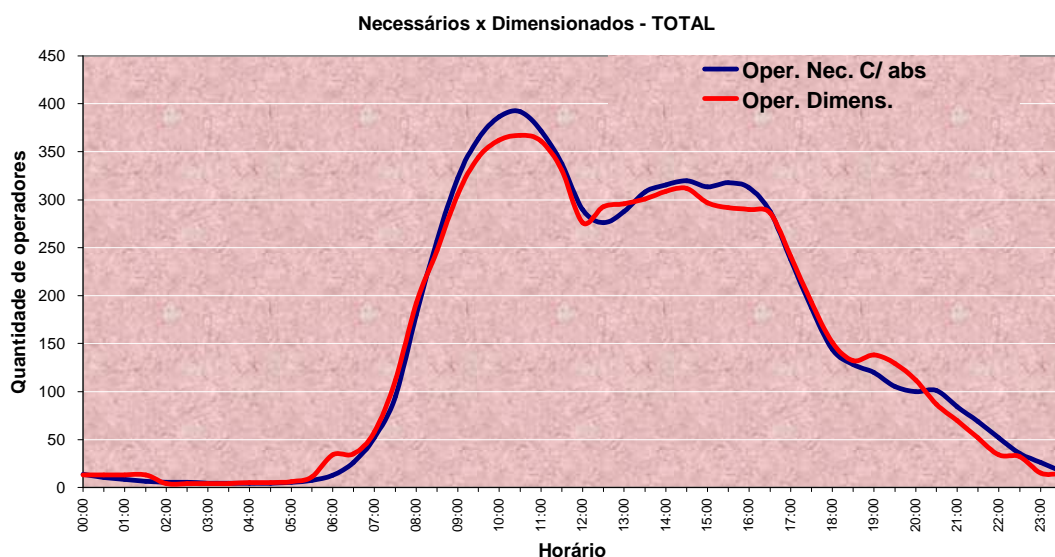


Fonte: Contax

O resultado deste processo é uma previsão de demanda de ligações (volume e tempo médio de atendimento) para cada meia-hora. Com o auxílio de fórmulas Erlang<sup>i</sup> do suplemento de Excel chamado Turbotab, é calculada a quantidade necessária de operadores para atender a demanda de cada período com um nível mínimo de serviço pré-estabelecido (normalmente 85% das ligações sendo atendidas em até 10 segundos).

O contingente atual (utilizado no mês anterior) de atendentes é, então, considerado. Em função da quantidade de operadores que estão iniciando a sua jornada a cada período do dia e da carga diária de trabalho de cada um (4 ou 6 horas), uma planilha calcula quantos atendentes estarão disponíveis em cada intervalo de meia-hora. Essas informações são, então, cruzadas com a necessidade de operadores para cada período de 30 minutos, levantada anteriormente. Tal cruzamento é sumarizado de maneira gráfica, como exemplificado na figura 2 a seguir.

**Figura 2 – Necessidade e disponibilidade de agentes por período, ago/06**



Fonte: Contax

Em cima do atual escalonamento é que a equipe de planejamento vai trabalhar para modificar a disponibilidade de operadores em cada período do dia, de modo a atingir o nível de serviço desejado. O objetivo é garantir que haja a quantidade certa de pessoas em cada horário, através do processo de tentativa-e-erro, durante o qual é preciso analisar diversos fatores, tais como: carga horária diária, aspectos da legislação trabalhista, acordos sindicais, espaço físico disponível. No caso do produto 102, o escalonamento balanceado (alternando momentos com contingente operacional abaixo/acima do necessário) pode ser usado, já que o importante para efeitos comerciais é o nível de serviço médio do dia.

Durante o processo de escalonamento, a equipe de planejamento faz tentativas modificando a quantidade de operadores que iniciam a sua jornada a cada horário. Essas mudanças alteram, consequentemente, a quantidade disponível de atendentes em cada intervalo de meia-hora. A planilha com as fórmulas Erlang utiliza, então, essa informação para estimar o nível de serviço para cada período de meia-hora e para o dia, em função da demanda prevista.

Neste processo iterativo, a principal motivação do analista é maximizar o nível de serviço médio do dia. O nível de serviço em cada faixa de horário, em si, não costuma ser



uma grande preocupação para o analista que, no entanto, procura evitar grandes déficits de atendentes escalados em relação aos demandados nas faixas de horários ao longo do dia.

A preocupação com os déficits intradiários existe porque, nos horários com muita carência de atendentes, pode haver grande incidência de abandonos. E isso pode ser muito ruim por dois motivos: a multa por excesso de ligações abandonadas e a possibilidade de o cliente desistente retornar posteriormente a ligação e aguardar até ser atendido, deteriorando o nível de serviço.

Esse esforço de dimensionamento e escalonamento objetiva um melhor ajuste entre a capacidade demandada e ofertada, ao longo do dia. Um exemplo das modificações realizadas na busca do melhor dimensionamento pode ser visualizado na figura 3 a seguir.

**Figura 3 – Mudanças realizadas durante o escalonamento dos atendentes, ago/06<sup>ii</sup>**

Horário	06:00	06:30	07:00	07:30	08:00	08:30	09:00	09:30	10:00	10:30
02:30										
03:00					05 Recursos de Básico					
03:30										
04:00										
04:30										
05:00										
05:30										
06:00	20									
06:30	20	1								
07:00	20	1	21							
07:30	20	1	21	25						
08:00	15	1	21	25	72					
08:30	17	0	15	25	72	40				
09:00	18	1	20	21	72	40	40			
09:30	20	1	21	24	72	40	40	26		
10:00	20	1	21	24	72	40	40	26	18	
10:30	20	1	21	25	72	40	40	26	18	9

Fonte: Contax

Na última etapa do processo de dimensionamento e escalonamento, o analista procura estimar como será o nível de serviço da operação (percentual de chamadas atendidas em até 10 segundos), para todos os dias ao longo do mês (até o momento, os cálculos foram feitos em cima do quinto dia de maior movimento apenas). Para tal, a distribuição intradiária de atendentes elaborada durante os passos anteriores é replicada para todos os dias do mês e, em conjunto com a previsão diária de demanda de ligações e com o perfil de comportamento intradiário da demanda, é capaz de estimar – através da mesma metodologia Erlang – os níveis de serviço a serem obtidos para cada dia e horário, dentro do mês em questão.

Na opinião da equipe de planejamento, a fórmula Erlang utilizada para o cálculo do nível de serviço não é totalmente precisa, mas não é radicalmente imprecisa a ponto de chamar a atenção e gerar situações onde o nível de serviço real se distancie muito do calculado. Na verdade, alguns questionamentos internos surgiram em relação a essa fórmula a partir de algumas observações empíricas, como em uma situação em que o nível de serviço calculado para um horário com déficit de 3 agentes era 77% e passava para 0% quando o déficit subia para 12 agentes. Dessa forma, não existe concordância de que essa metodologia é a ideal para calcular o nível de serviço, mas o setor não encontrou nenhuma outra abordagem mais precisa, em pesquisas coletando informações e verificando planilhas.

### 3.3. Metodologia sugerida para o dimensionamento da capacidade de atendimento

Para a realidade dos *call centers*, a Teoria de Filas pode ser a melhor metodologia analítica a ser utilizada, mas existem métodos experimentais – como a simulação – que podem ser mais adequados para uma indústria com o cotidiano operacional tão complexo quanto o de

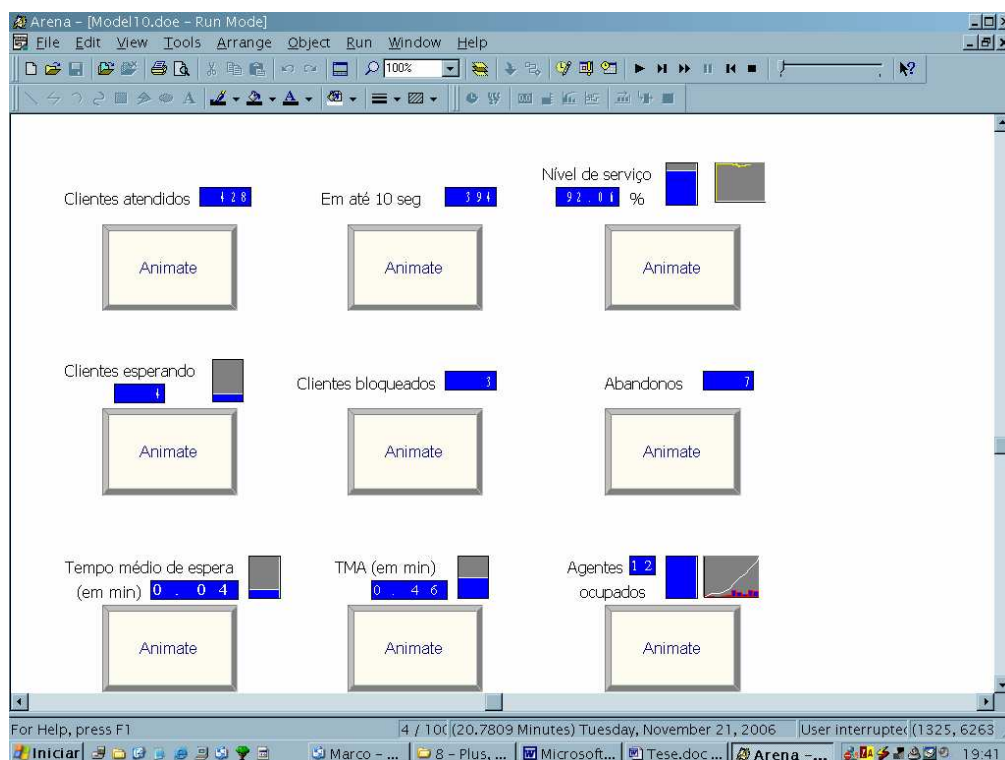
um *call center* moderno, conforme sugerido na seção 2.2 deste trabalho. Em especial, a metodologia atualmente utilizada para o cálculo do nível de serviço não considera a hipótese de abandono por parte do cliente ou sinal de ocupado; ou seja, os clientes que chegam e não encontram agentes disponíveis, aguardam indefinidamente em uma fila até serem atendidos. Isso cria uma tendência de subestimar o nível de serviço porque, na operação real, alguns clientes vão abandonar as ligações (se esperarem muito), diminuindo a fila e fazendo com que sejam atendidas em menos tempo as chamadas programadas para serem respondidas após aquela que abandonou a fila. Adicionalmente, as planilhas usadas para o cálculo do nível de serviço consideram que o tempo de atendimento segue uma distribuição exponencial. No entanto, isso raramente ocorre na prática, caracterizando uma premissa relativamente forçada.

O emprego da simulação permite contemplar as características destacadas na seção 2.2, incluindo o comportamento de abandono (podendo também considerar que um percentual dos clientes que abandonam irá retornar a ligação, tentando um novo contato dentro de uma quantidade de tempo, que pode ser modelada por uma distribuição estatística) e uma flexibilidade na definição do formato da distribuição do tempo de atendimento.

Para verificar essa melhor adequação, a idéia consiste em simular computacionalmente – em poucos segundos – a operação do *call center* durante períodos de 30 minutos, contemplando de forma mais realista as suas características para obter resultados mais acurados do que com as metodologias analíticas.

Dessa forma, é possível visualizar a operação em si (com as chamadas chegando, sendo enfileiradas e depois atendidas) e o que estaria acontecendo, de forma detalhada (praticamente como se fosse *in loco*), para poder entender porque determinado período do dia apresentou um nível de serviço tão baixo, por exemplo (ao invés de apenas aceitar o número frio fornecido pelas fórmulas analíticas). A figura 4 a seguir ilustra a apresentação dinâmica dos indicadores de desempenho que um *software* de simulação é capaz de oferecer.

**Figura 4 – Indicadores dinâmicos de um modelo de simulação para *call center***



Fonte: Tela capturada durante a simulação do modelo pelo *software*

Como pode ser visto, diversos indicadores podem ser acompanhados – à medida que vão variando dinamicamente enquanto a simulação vai sendo executada: a quantidade de clientes atendidos, atendidos em até 10 segundos, o consequente nível de serviço (seu valor no momento, na forma numérica e gráfica – a barra de nível, além do seu comportamento ao longo da simulação), o número de clientes esperando na fila (na forma numérica e gráfica), de clientes bloqueados e que abandonaram a fila, os tempos médios de espera e de atendimento (na forma numérica e gráfica), e a quantidade de agentes ocupados (seu valor no momento, na forma numérica e gráfica, além de um histograma registrando seu comportamento ao longo da simulação). No final da simulação, os relatórios sumarizam os resultados consolidados.

Com o uso da simulação, é possível calcular empiricamente (ao invés de estimar analiticamente) – através do uso de algumas premissas históricas (de previsão de demanda e comportamento dos clientes e do sistema) e da quantidade dimensionada de atendentes como *inputs* – alguns indicadores de performance importantes, como: nível de serviço, taxa de abandono, tempo médio de espera, utilização e ociosidade dos agentes.

Para o dimensionamento e escalonamento dos operadores para atender os clientes do produto 102 durante o mês de agosto de 2006, foi usada a premissa (oriunda da previsão de demanda) de que chegariam à central 586 chamadas, com um tempo médio de atendimento de 29 segundos, na primeira meia-hora do dia (00:00 às 00:30). Seria necessário, segundo as fórmulas analíticas, alocar 12 agentes para este horário, de forma a atingir 85% de nível de serviço. A equipe de dimensionamento alocou, então, 12 atendentes e as fórmulas analíticas previram 88,04% de nível de serviço para este período.

Com o objetivo de questionar esses números, foi construído – no *software* Arena Contact Center – um modelo para simular como se comportaria o sistema nesse horário, com as mesmas premissas de demanda (volume e tempo médio de atendimento) e com a mesma capacidade operacional (12 agentes).

Como as chamadas vão chegando à central sem nenhuma espécie de controle, este processo pode ser considerado aleatório, caso em que a base conceitual sugere que a taxa de chegada de ligações seja modelada através de um processo de Poisson. O modelo de simulação construído implementou este processo com média de, aproximadamente, 0,33 ligações chegando por segundo (ou 586 em um intervalo de 30 minutos).

Em relação ao tempo de atendimento, a distribuição Erlang costuma modelar bem esse processo e, portanto, foi utilizada, com média de 29 segundos. Ela requer, no entanto, um parâmetro adicional ( $k$ ) que diz respeito à variabilidade dos dados em torno da média. O desvio-padrão da distribuição é igual à sua média dividida pela raiz quadrada de  $k$ . Para, então, considerar uma variabilidade moderada dos dados em torno da média, o modelo utiliza a distribuição Erlang com  $k = 4$ , fazendo com que o coeficiente de variação seja de 50%.

Para que o modelo contemplasse corretamente o comportamento de abandono dos clientes, foi necessário realizar uma pesquisa junto à base de dados da Contax, contendo as ligações abandonadas, para o produto 102. Ela mostrou que o tempo de espera das ligações abandonadas apresenta, historicamente, uma média em torno de 2,5 minutos, seguindo uma distribuição não muito distante de uma exponencial. Também foi preciso modelar o comportamento de retorno das ligações abandonadas. Para tal, foi utilizada a premissa de que 80% das ligações abandonadas são refeitas posteriormente, entre 1 e 9 minutos após o abandono (distribuição uniforme).

### 3.4. Análise dos resultados

A simulação dos 30 minutos de operação do *call center* foi replicada 100 vezes no *software* em 142 segundos e o primeiro resultado indica que, em média, 595 chamadas foram geradas em cada replicação. Este número é um pouco maior do que a premissa de demanda

(586 chamadas) porque, na simulação, algumas das chamadas abandonadas voltaram a ser geradas e entraram novamente na fila. Das chamadas geradas, 579 – em média – foram efetivamente atendidas pelos agentes em cada replicação.

Destas chamadas, 541 foram atendidas em até 10 segundos, encerrando um nível de serviço de 93,31%. Esse valor é razoavelmente maior do que os 88,04% previstos de nível de serviço segundo a abordagem analítica. Esse fato confirma, em princípio, a expectativa de subestimação desta grandeza por parte da Teoria de Filas.

Depois de realizada uma consulta junto à base de dados da Contax, foi possível recuperar a informação de que, no dia 22 de agosto de 2006, estiveram presentes 12 operadores durante o período das 00:00 às 00:30. Dentro deste intervalo, 592 ligações foram atendidas, em 29,4 segundos cada, em média. Estes números superam por muito pouco as premissas de demanda (volume = 586; tempo média de atendimento = 29 segundos) utilizadas no dimensionamento via fórmulas analíticas e também no modelo de simulação aqui apresentado. Por isso, esse dia e horário foram escolhidos para servirem de base de comparação entre os resultados obtidos pelas duas abordagens.

Durante estes 30 minutos, 549 das 592 chamadas foram atendidas em até 10 segundos; ou seja, o nível de serviço real ocorrido neste intervalo foi de 92,74%, bem mais próximo do valor calculado empiricamente pela simulação (93,31%) do que do valor estimado analiticamente pelas fórmulas Erlang (88,04%). Conforme foi dito, a subestimação do nível de serviço promovida pela Teoria de Filas deve-se, sobretudo, à não contemplação do abandono da ligação, empreendido – na prática – por uma parcela dos clientes, mas desconsiderado – na teoria – pelos modelos analíticos.

Das 595 chamadas geradas em cada replicação, 14,5 – em média – foram abandonadas pelos clientes, ocasionando uma taxa de abandono igual a 2,44%. Dentre as 14,5 chamadas abandonadas, 11,5 (79,41%) retornaram para a fila alguns minutos após o abandono.

Como deve estar sendo observado, a simulação permite que muitos outros indicadores de performance – além do nível de serviço – a respeito das chamadas possam ser observados e analisados. Por exemplo, o tempo médio de atendimento foi de 29,35 segundos. Mas um cliente chegou a ser atendido em 4,70 segundos e outro em 147,60 segundos. Já em relação ao tempo de espera, as chamadas aguardaram, em média, 1,94 segundos para serem atendidas. Alguns clientes foram atendidos instantaneamente e pelo menos um esperou quase 38 segundos antes de ser atendido.

Conforme alertado na seção 2.2, esse tipo de análise envolvendo a variabilidade e os valores máximos e mínimos dos indicadores de desempenho não pode ser realizado por meio de métodos analíticos (capazes de apresentar apenas os valores médios), tornando-se viável apenas através de abordagens experimentais.

Naturalmente, a simulação também gera indicadores a respeito dos agentes. No período em questão, a taxa média de ocupação dos atendentes foi de 78,75%. Este indicador pode orientar a gerência no sentido de um aumento ou redução de efetivo, de acordo com as metas de ocupação traçadas *a priori*.

É interessante observar também a acurácia da abordagem experimental em outros horários, com diferentes patamares para o nível de serviço.

Para o dimensionamento dos operadores (também para agosto de 2006) no período entre 02:00 e 02:30, foi usada a premissa de que chegariam à central 196 chamadas, com um tempo médio de 28 segundos. Seria necessário, segundo as fórmulas analíticas, alocar 5 agentes para este horário, de forma a atingir 85% de nível de serviço. Foram dimensionados, no entanto, 4 atendentes para este período e este déficit de agentes (-1) fez as fórmulas analíticas preverem 44,05% de nível de serviço para este horário.

Novamente foi construído um modelo, desta vez para simular o comportamento do sistema neste segundo horário, com as mesmas premissas de demanda (volume e tempo médio de atendimento) e com a mesma capacidade dimensionada (4 agentes). A chegada das ligações e o tempo de atendimento foram modelados de acordo com estas premissas médias e seguindo as distribuições exponencial e Erlang ( $k = 4$ ), respectivamente; e o comportamento de abandono modelado da mesma forma que anteriormente.

Em média, 209 chamadas foram geradas em cada replicação; destas, 193 – em média – foram de fato atendidas pelos operadores em cada replicação, das quais 144 em até 10 segundos. O nível de serviço obtido foi 74,44%, um valor mais baixo que a meta de 85%, mas muito acima dos 44,05% previstos pelas fórmulas analíticas que, de fato, parece ter subestimado esta grandeza.

Segundo a base de dados da Contax, no dia 29 de agosto de 2006, 4 operadores atenderam 192 ligações, entre 02:00 e 02:30, com tempo médio de atendimento de 27,6 segundos. Estes números são bem próximos das premissas de demanda utilizadas no dimensionamento através de fórmulas Erlang e no de simulação. Nesse intervalo, 136 ligações foram atendidas em até 10 segundos, gerando um nível de serviço real de 70,83%.

Assim como quando aconteceu no cenário original (horário das 00:00 às 00:30), este valor ficou mais próximo (desta vez ainda mais) do nível de serviço calculado pela simulação. Isso tudo revela um ganho ainda maior de acurácia na previsão do nível de serviço, em cenários com baixos patamares para essa grandeza. Resta saber o comportamento desta acurácia em situações com níveis de serviço muito altos (saldo de agentes dimensionados em relação à necessidade).

No horário entre 05:30 e 06:00, durante o mês de agosto de 2006, foi considerado que chegariam à central 253 chamadas, com um tempo médio de atendimento de 31 segundos. Segundo as fórmulas analíticas, seria necessário designar 7 atendentes para este período, de forma a alcançar a meta de nível de serviço. Foram dimensionados 11 agentes (saldo de 4) para este horário e as fórmulas Erlang previram 99,78% de nível de serviço para este horário.

Com o objetivo de questionar esse nível de serviço estimado, foi construído um modelo com as mesmas premissas de demanda (volume e tempo médio de atendimento) e com a mesma capacidade dimensionada (11 atendentes), desta vez para este terceiro horário.

A chegada das ligações e o tempo de atendimento foram modelados de acordo com estas premissas médias e seguindo as distribuições exponencial e Erlang ( $k = 4$ ), respectivamente; e o abandono foi tratado de forma idêntica à anterior.

A simulação da operação durante este intervalo foi replicada 100 vezes. Em média, 253,31 chamadas foram geradas, 253,27 atendidas e 253,22 em até 10 segundos em cada replicação. Quase não houve abandono, o que não é surpreendente em vista do superdimensionamento empreendido para este cenário. O nível de serviço resultante foi 99,98%, um valor altíssimo (inclusive quando comparado com a meta de 85%), e bem parecido com os 99,78% previstos pelas fórmulas Erlang para este cenário.

A base de dados da Contax revela que, no dia 15 de agosto de 2006, 11 agentes estiveram atendendo durante o período das 05:30 às 06:00, quando 249 chamadas foram atendidas, com tempo médio de 31,5 segundos. Estes números estão muito próximos das premissas de demanda usadas nos modelos analítico e de simulação.

Dentro de tal período, foram atendidas 248 chamadas em até 10 segundos, gerando um nível de serviço real de 99,60%. Desta vez, os valores previstos por ambas as abordagens (analítica e experimental) ficaram muito parecidos com o valor real, até porque não havia muito espaço para grandes variações, visto que este cenário envolve um patamar de nível de serviço já muito próximo de 100%. A conclusão aponta para a inexistência de ganho de



acurácia na previsão do nível de serviço – em cenários com patamares muito altos para essa grandeza – promovido pelo uso da abordagem experimental.

Finalmente é possível apresentar uma comparação mais completa. O comportamento do ganho de acurácia – empreendido pelo uso da abordagem de simulação – em função do patamar onde se encontra o nível de serviço pode ser resumido na tabela 1 a seguir.

**Tabela 1 – Nível de serviço real e estimado por fórmulas Erlang e simulação, erros de estimação e ganho de acurácia em diferentes patamares de nível de serviço**

Horário	02:00 às 02:30	00:00 às 00:30	05:30 às 06:00
Saldo/déficit de agentes	1 a menos em 5 -20%	0 em 12 0%	4 a mais em 7 +57%
Nível de serviço real	70,83%	92,74%	99,60%
Fórmula Erlang	44,05%	88,04%	99,78%
Erro	26,78%	4,70%	0,18%
Simulação	74,44%	93,31%	99,98%
Erro	3,61%	0,57%	0,38%
Diferença entre os erros (ganho de acurácia)	23,17%	4,13%	-0,20%

Fonte: Tabela elaborada pelo autor

#### 4. Conclusões

Foi possível constatar, ao longo dessa pesquisa e através do uso de modelos de simulação para lidar com o problema de dimensionamento da capacidade de atendimento do *call center* da empresa estudada, algumas vantagens da abordagem experimental em comparação às analíticas, especialmente em operações mais complexas: (i) é possível incluir mais detalhes da operação, utilizar distribuições estatísticas mais condizentes com os dados de entrada e deixar o modelo mais próximo da realidade, garantindo a obtenção de resultados mais acurados; (ii) o nível de serviço calculado pelas fórmulas Erlang normalmente é subestimado, principalmente pelo fato destas ignorarem o abandono de chamadas; (iii) outros indicadores de desempenho (não disponíveis nas abordagens analíticas, como a taxa de abandono) podem ser avaliados, apresentados e estudados; (iv) valores mínimo e máximo de cada indicador importante podem ser obtidos, não ficando o analista restrito à média dos valores, como no caso da Teoria de Filas; (v) um entendimento melhor da operação é atingido com a adoção da abordagem experimental, onde existe a possibilidade de acompanhar dinamicamente o comportamento do sistema e dos seus indicadores de performance e, dessa forma, entender por que as filas estão se formando e o motivo dos tempos de espera estarem altos, por exemplo, ao passo que, através da metodologia Erlang, é possível enxergar apenas os *outputs* gerados (indicadores numéricos) em função dos *inputs* fornecidos, dificultando o entendimento completo da operação; (vi) a comunicação pode tornar-se mais fácil, através de animações gráficas.

O estudo em questão também permitiu concluir que o ganho de acurácia para o nível de serviço promovido pela simulação tende a ser maior nos horários onde essa grandeza se encontra em patamares mais baixos, conforme a tabela 1 anterior.

##### 4.1. Sugestões e recomendações

Muito tem sido feito desde que se percebeu a complexidade inerente ao processo de gerenciamento de *call centers* modernos. Naturalmente, há ainda muito espaço para serem respondidas – mais e de forma mais acurada – questões relativas à otimização de centrais de atendimento, através de mais discussões e novos estudos, e da melhoria em um aspecto fundamental nos sistemas de modelagem: a proximidade com a realidade.

Um objeto de pesquisa interessante seria o mapeamento – e a inclusão no modelo – do comportamento pós-abandono dos clientes: a parcela percentual que tenta ligar novamente e o formato da distribuição do tempo transcorrido até a próxima tentativa. Adicionalmente, poderia também ser levado em consideração o fato de que, após abandonar a ligação inicialmente, o cliente que volta a tentar contato pode estar mais impaciente e acabar aguardando menos tempo desta vez, antes de abandonar novamente.

Finalmente, trabalhos futuros de simulação na indústria de *call centers* poderiam explorar outras peculiaridades deste tipo de operação, que não costumam ser bem abordadas através das metodologias analíticas, como, por exemplo: (i) o processo de transferência da chamada durante o atendimento a um cliente, até que este seja servido pelo funcionário correto; (ii) conferências entre o cliente e mais de um atendente ao mesmo tempo; (iii) desvios condicionais de chamadas para atendimentos especializados; e (iv) outras disciplinas de fila, diferentes da tradicional FIFO.

Já que a maioria dos *softwares* de simulação contempla tais atributos operacionais, não seria complicado – para um pesquisador interessado nestas sugestões e disposto a se aprofundar nos detalhes funcionais da ferramenta por ele utilizada – modelar estas peculiaridades e alcançar conclusões interessantes para a indústria em questão.

### **Referências Bibliográficas**

- ALAM, M. Using Call Centers to Deliver Public Services. *House of Commons Paper*. London: The Stationery Office Books, 2002
- ARAÚJO, M.; ARAÚJO, F.; ADISSI, P. Modelo para segmentação da demanda de um call center em múltiplas prioridades: estudo da implantação em um call center de telecomunicações. *Revista Produção On Line*, volume 4(3), Agosto 2004
- AVRAMIDIS, A.; L'ECUYER, P. Modeling and Simulation of Call Centers. *Winter Simulation Conference*, 2005
- BAPAT, V.; PRUITTE Jr, E. Using simulation in call centers. *Winter Simulation Conference*, p. 1395-1399, 1998
- BATT, R.; DOELLGAST, V.; KWON, H. U. S. Call Center Industry Report 2004 – National Benchmarking Report – Strategy, HR Practices & Performance. *Center for Advanced Human Resource Studies (Cornell University)*, Working Paper 05 – 06, 2004
- BOUZADA, M. *O uso de ferramentas quantitativas em call centers: o caso Contax*. Orientador: Eduardo Saliby. Rio de Janeiro: UFRJ/COPPEAD, 2006. Tese. (Doutorado em Administração de Empresas)
- BROWN, L.; GANS, N.; MANDELBAUM, A.; SAKOV, A.; SHEN, H.; ZELTYIN, S.; ZHAO, L. Statistical analysis of a telephone call center: a queueing-science perspective. *Wharton Financial Institutions Center*, Working Paper, 03 – 12, Nov 9, 2002
- CHASSIOTI, E.; WORTHINGTON, D. A new model for call centre queue management. *Journal of the Operational Research Society*, vol. 55, p. 1352-1357, 2004
- CHOKSHI, R. Decision support for call center management using simulation. *Winter Simulation Conference*, p. 1634-1639, 1999
- CONTAX. *Contax Contact Center*. Disponível em: [www.contax.net.br](http://www.contax.net.br). Acesso em: 19 jul. 2006
- GROSSMAN, T.; SAMUELSON, D.; OH, S.; ROHLER, T. Call Centers. *Encyclopedia of Operations Research and Management Science*, Boston: Kluwer Academic Publishers, p. 73-76, 2001
- GULATI, S.; MALCOLM, S. Call center scheduling technology evaluation using simulation. *Winter Simulation Conference*, 2001

- HALL, B.; ANTON, J. Optimizing your call center through simulation. *Call Center Solutions Magazine*, Oct. 1998
- HAWKINS, L.; MEIER, T.; NAINIS, W.; JAMES, H. Planning Guidance Document For US Call Centers. *Information Technology Support Center*. Maryland, 2001
- HERTZ, D. Análise de risco em investimentos de capital. *Biblioteca Harvard de Administração de Empresas*, 8(3), p. 1-14, 1980
- HILLIER, F.; LIEBERMAN, G. *Introduction to Operations Research*. New York: McGraw-Hill, 1995
- KLUNGLE, R. Simulation of a claims call center: a success and a failure. *Winter Simulation Conference*, p. 1648-1653, 1999
- KLUNGLE, R.; MALUCHNIK, J. The role of simulation in call center management. *MSUG Conference*, 1997
- MEHROTRA, V. Ringing Up Big Business. *OR/MS Today*, 24(4), p.18-24, Aug 1997
- MEHROTRA, V.; FAMA, J. Call Center Simulation Modeling: Methods, Challenges and Opportunities. *Winter Simulation Conference*, p. 135-143, 2003
- MEHROTRA, V.; PROFOZICH, D.; BAPAT, V. Simulation: the best way to design your call center. *Telemarketing & Call Center Solutions*, Nov 1997
- OUTSOURCING. *Ranking*. Disponível em: [www.callcenter.inf.br](http://www.callcenter.inf.br). Acesso em: 20 set. 2005
- PARAGON. *Simulação de Call Center com Arena Contact Center*. Disponível em: [www.paragon.com.br](http://www.paragon.com.br). Acesso em: 20 set. 2005
- PIDD, M. *Computer Simulation in Management Science*. New York: Willey, 1998
- RAGSDALE, C. *Spreadsheet Modeling and Decision Analysis*. Tennessee: South-Western, 2001
- RILEY, D. Simulating a Virtual Customer Service Center. *Winter Simulation Conference*, 2005
- SALIBY, E. *Repensando a Simulação: a Amostragem Descritiva*. São Paulo: Atlas, 1989
- STECKLEY, S.; HENDERSON, S.; MEHROTRA, V. Performance Measures for Service Systems with a Random Arrival Rate. *Winter Simulation Conference*, 2005
- TANIR, O.; BOOTH, R. Call center simulation in Bell Canada. *Winter Simulation Conference*, p. 1640-1647, 1999
- WEINBERG, J.; BROWN, L.; STROUD, J. Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data. *University of Pennsylvania*, White Paper, Jun 2006
- WORTHINGTON, D.; WALL, A. Using the discrete time modelling approach to evaluate the time-dependent behaviour of queueing systems. *Journal of the Operational Research Society*, vol. 50, p. 777-788, 1999
- YONAMINE, J. *O Setor de Call Centers e Métodos Quantitativos: uma Aplicação da Simulação*. Orientador: Eduardo Saliby. Rio de Janeiro: UFRJ/COPPEAD, 2006. Dissertação. (Mestrado em Administração de Empresas)

---

<sup>i</sup> A Teoria de Filas é utilizada para o dimensionamento de agentes na Contax. Os clientes que chegam e não encontram atendentes disponíveis, aguardam indefinidamente em uma fila até serem atendidos (não existe abandono ou sinal de ocupado), segundo a teoria. Os tempos de espera são previstos partindo da premissa que o intervalo entre chegadas e o tempo de atendimento seguem distribuições exponenciais, e levando em consideração três fatores: o número de agentes; o número de clientes em espera na fila; e o tempo médio de atendimento.

<sup>ii</sup> As quantidades de atendentes disponíveis destacadas em azul são menores do que as de operadores que iniciam a jornada em cada período por causa da ocorrência do horário de lanche para parte dos funcionários nestes momentos.