

Hat der Servicelevel ausgedient?

Genügt der Servicelevel als einzige Kennzahl zur Personaleinsatzplanung in Call Centern?

Der dominierende Parameter der Kapazitätsplanung von Inbound-Call Centern ist der Servicelevel. Andere Größen spielen eine untergeordnete Rolle. Liegt hier die Ursache dafür, dass viele Call Center offensichtlich nicht richtig in den Griff zu bekommen sind?

von Florian Schümann

Zweck der Kapazitätsplanung in Inbound-Call Centern ist es, die Erreichbarkeit an den Erwartungen der Zielgruppe auszurichten. Die Planung basiert mangels Alternativen überwiegend auf der Warteschleifentheorie, die als 'Erlang C' in der Branche geläufig ist. Darum ist der Servicelevel die wohl prominenteste Call Center-Kennziffer.

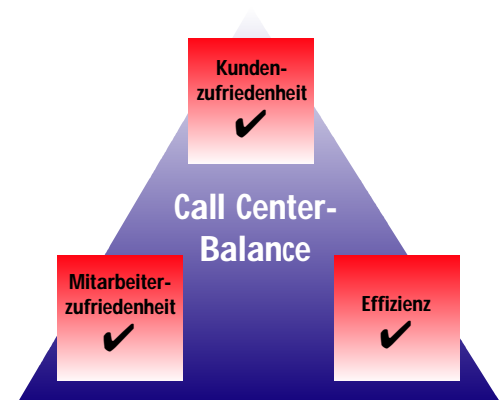
Da Call Center natürlich ökonomisch sinnvoll agieren müssen und in Zeiten hoher

Fluktuation und mangelnder Akzeptanz die Mitarbeiterzufriedenheit eine immer gewichtigere Rolle spielt, ist es an der Zeit, sich von der 'servicelevelfixierten' Planung zu verabschieden und gleichberechtigt einen 'Effizienzlevel' und einen 'Mitarbeiterlevel' in die Planungskonzeption zu integrieren.

Zielkonflikte

Kunden klagen über Warteschleifen, Mitarbeiter kommen und gehen, und unternehmensintern gelten Call Center oft als Kostentreiber. Plant man mehr Personal ein, um die Kunden zufriedenzustellen, beschwert sich das Controlling über zu hohe Personalkosten. Reduziert man den Personaleinsatz, indem man 'mehr aus den Mitarbeitern herausholt', freuen sich zwar Controller und Kunden, aber die Mitarbeiter klagen.

Offensichtlich stehen die Ziele in Konkurrenz zueinander. Eine synchrone Optimierung ist schwer zu realisieren. Denn schon die zur Verfügung stehenden Instrumentarien, also die (modifizierten) Erlang-C-Modelle sind nicht darauf ausgelegt, mehrere Ziele gleichzeitig zu optimieren. Die Warteschleifentheorie wurde dafür entwickelt, die Anzahl von Bedienkräften zu ermitteln, die einen



Ziel der Kapazitätsplanung im Call Center ist es, die Balance der drei Komponenten herzustellen.

bestimmten Arbeitsanfall auf einem bestimmten Serviceniveau bewerkstelligen kann. Bezüglich der beiden anderen Ziele werden lediglich mathematische Tatsachen geschaffen. Auf den jeweiligen Zielerreichungsgrad wird nicht reflektiert.

Eine plausible Anzahl von Bedienkräften zu errechnen, ist zunächst einfach (siehe Kasten 'Kalkulation der Mitarbeiter'). Komplizierter ist es, auf ein ökonomisch sinnvolles Ergebnis hinzuplanen. Oft begnügt man sich damit, das Jahres- oder Monatsbudget oder eine Personalkostenreduktion von fünf Prozent im Blickwinkel zu haben. Beliebt ist es auch,

Beispiel

1. Kalkulation der Mitarbeiter

Gesucht wird die Anzahl der benötigten Mitarbeiter in einem 30-Minuten-Intervall. Der Servicelevel wird mit 80 Prozent in 20 Sekunden vorgegeben, die durchschnittliche Bearbeitungsdauer ist mit 100 Sekunden ermittelt worden, 1000 Anrufe sind prognostiziert worden. Nach Erlang C ergibt sich folgendes Szenario, wobei die Kosten pro Call vorgegeben werden:

Im Ergebnis werden 60 Mitarbeiter benötigt, um einen Servicelevel von mindestens 80 Prozent zu erreichen.

2. Effizienz

Kosten pro Call von 0,90 DM seien durch Erlöse pro Call von 0,90 DM abgedeckt. Die Mitarbeiterauslastung errechnet sich mit 93 Prozent (27,9 von 30 Minuten verbringen die Agenten mit Telefonieren und Nacharbeiten). Zwei von drei Zielen wären damit erreicht.

Benötigte Agenten	durchschnittliche Zeit bis Anrufannahme (ASA) (in s)	Servicelevel (in Prozent)	Agentenauslastung (in Prozent)	Kosten pro Call (in DM)
58	26	59	96	0,87
59	15	72	94	0,89
60	10	81	93	0,90
61	6	87	91	0,92
62	4	91	90	0,93

3. Auslastung

Entspricht die mathematisch korrekt ermittelte Anzahl von 60 Mitarbeitern (und damit korrespondierende Auslastung von 93 Prozent) aber den Erwartungen der Mitarbeiter? Wahrscheinlich nicht, denn 93 Prozent ist als hoch einzustufen. Zwei Situationen sind deshalb möglich:

Die errechnete Auslastung der Mitarbeiter ist entweder

- zu hoch (hier wahrscheinlich) oder
- zu niedrig.

Beide, sowohl eine zu hohe, als auch eine zu geringe Auslastung, werden von Call Cen-

ter Agents in der Regel als unbefriedigend empfunden.

Bei zu hoher Auslastung entwickeln Mitarbeiter das Gefühl, mehr Unterstützung durch zusätzliche Personalkapazitäten zu benötigen. Die zunächst berechnete Personalmenge wäre damit hinfällig, weil mehr Personal notwendig gewesen wäre. Das wiederum würde aber das ursprüngliche, partielle Gleichgewicht 'Servicelevel/Effizienz' kippen.

Die Reduktion der Auslastung durch mehr Mitarbeiterinsatz wäre die Folge. Wird sie zum Beispiel auf 90 Prozent verringert, wür-

alle Kosten durch alle Anrufe zu teilen und zu hoffen, mit so wenig Personaleinsatz wie möglich das produzierte Gut – den Anruf oder Prozess – so kostengünstig wie nötig herzustellen (siehe Kasten 'Effizienz').

Am schwierigsten ist es, Mitarbeiterbe-
lange, also Ein-
stellungen oder

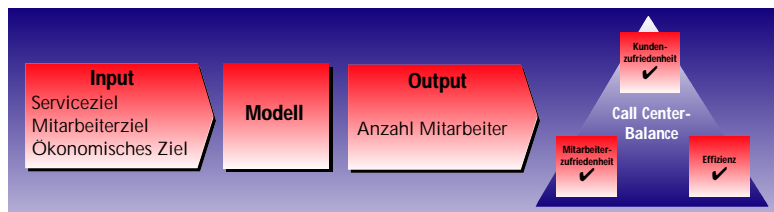
Erwartungen des Personals, mit zu berücksichtigen (siehe Kasten 'Auslastung'). Daran hat bei der Konzeption der Warteschleifenmodelle niemand gedacht. Zielverfehlungen in diesem Bereich wirken sich aber besonders schwerwiegend aus, zum Beispiel auf die Qualität des Produktes 'Anruf'.

Die Ursache dafür, dass der Balanceakt 'Kundenzufriedenheit-Effizienz-Mitarbeiterzufriedenheit' so schwer gelingt, ist vermutlich in der Warteschleifentheorie selbst zu finden (siehe Kasten 'Fazit').

Was ist zu tun?

Benötigt wird die Möglichkeit der Synchronplanung und -optimierung. Ansätze hierzu existieren bereits: Balanced Scorecards oder Simulatoren liefern eine gute Ausgangsbasis. Das allein reicht aber nicht. Es fehlt ein Modell – besser: eine Formel, die als Input den gewünschten Service-, Effizienz-

und Mitarbeiterlevel aufnimmt und als Output die Anzahl von Mitarbeitern ergibt, die einen ausgewogenen Zielerreichungsgrad aller Ziele liefert (siehe Grafik).



Wirkungszusammenhänge zum Erreichen der Call Center-Balance.

Ergo darf sich die Kapazitätsplanung nicht darauf beschränken, nach Servicelevel zu optimieren. Wenn ein Call Center ausbalanciert sein soll, muss sowohl ein 'Effizienzlevel' als auch ein 'Mitarbeiterlevel' gleichberechtigt in die Methodik integriert werden.

Entweder muss die Warteschleifentheorie um zwei Parameter erweitert oder ein alternatives Modell geschaffen werden. Beide Wege führen zwangsläufig dazu, die relative Bedeutung des Servicelevels zu reduzieren. Ausgedient hat er damit zwar nicht, aber Konkurrenz bekommt er allemal.

TT

Florian Schümann ist Call Center-Consultant bei InVisi - on Software GmbH in Ratingen. E-Mail: florian.schue-mann@invision.de



de das ursprüngliche Ungleichgewicht in ein neues Ungleichgewicht überführt. Der Servicelevel wäre 'zu hoch' – was aus Kundensicht zu ertragen wäre. Außerdem würde man den effizienten Bereich verlassen, da mehr Mitarbeiter eingesetzt werden müssten und damit die Kosten pro Call auf 0,93 DM anwachsen.

partielle Servicelevel/Effizienz-Gleichgewicht zu Lasten des Servicelevels stören.

4. Fazit des Beispiels

Das Beispiel zeigt, dass die Elemente des Zielsystems eng miteinander verbunden sind und Abhängigkeiten untereinander aufweisen. Einem Gleichgewichtszustand (SL=80/20, Auslastung 93%, Kosten/Anruf

Benötigte Agenten	durchschnittliche Zeit bis Anrufannahme (ASA) (in s)	Servicelevel (in Prozent)	Agentenauslastung (in Prozent)	Kosten pro Call (in DM)
58	26	59	96	0,87
59	15	72	94	0,89
60	10	81	93	0,90
61	6	87	91	0,92
62	4	91	90	0,93

Damit löst wiederum ein Ungleichgewicht das andere ab.

Im Falle einer als zu gering empfundenen Auslastung lassen sich analoge Überlegungen anstellen: Kommt das Gefühl von Unterlastung auf, hätte eine geringere Besetzung ausgereicht. Die Erhöhung der Auslastung würde hier in anderer Richtung das

0,90) stehen aufgrund der hohen Komplexität des Systems eine Vielzahl möglicher nicht optimaler Zustände gegenüber. Das Beispiel zeigt auch, dass das Optimieren nach Servicelevel allein unzureichend ist, dass die Kosten pro Call und die Auslastung als Brücke zu den Erwartungen lediglich Orientierung bieten. ■