



# BUNDESPATENTGERICHT

IM NAMEN DES VOLKES

URTEIL

Verkündet am  
6. März 2024

...

4 Ni 80/22

---

(Aktenzeichen)

In der Patentnichtigkeitssache

...

**betreffend das deutsche Patent 10 2013 001 219**

hat der 4. Senat (Nichtigkeitssenat) des Bundespatentgerichts auf die mündliche Verhandlung vom 6. März 2024 durch den Vorsitzenden Richter Voit und die Richter Dipl.-Ing. Altvater, Dr. von Hartz, Dipl.-Phys. Univ. Dr. Haupt sowie Dipl.-Ing. Tischler

für Recht erkannt:

- I. Die Klage wird abgewiesen.
- II. Die Klägerin trägt die Kosten des Rechtsstreits.
- III. Das Urteil ist im Kostenpunkt gegen Sicherheitsleistung in Höhe von 120 % des zu vollstreckenden Betrages vorläufig vollstreckbar.

**T a t b e s t a n d**

Die Beklagte ist Inhaberin des deutschen Patents 10 2013 001 219 (Streitpatent), das am 25. Januar 2013 angemeldet worden ist. Die Erteilung des Patents ist am 29. August 2019 veröffentlicht worden. Das Streitpatent ist in Kraft und trägt die Bezeichnung

„Verfahren und System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus“.

Es umfasst in der erteilten Fassung fünfzehn Patentansprüche, die die Klägerin mit ihrer Nichtigkeitsklage vom 10. Oktober 2022 im gesamten Umfang angreift.

Der das Verfahren betreffende unabhängige Patentanspruch 1 und der die korrespondierende Vorrichtung betreffende unabhängige Patentanspruch 12 lauten in der erteilten Fassung:

1. Verfahren zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus, **dadurch gekennzeichnet**,

a) dass Audio-Daten (11) mit mindestens einem Mikrofon (2) aufgenommen werden,

b) dass die Audio-Daten (11) kontinuierlich in mindestens einem Audio-Puffer (6) zwischengespeichert werden, so dass der Audio-Puffer (6) stets die Audio-Daten (11) der jüngsten Vergangenheit enthält,

c) dass die Audio-Daten (11) zeitnahe mindestens einem sekundären Spracherkennungs-Prozess (7) zugeführt werden, welcher beim Erkennen eines Schlüsselworts (18) oder einer Phrase mindestens einen primären Spracherkennungs-Prozess (8) startet oder aus einem Ruhezustand aktiviert,

d) dass der primäre Spracherkennungs-Prozess (8) den gesamten oder den jüngsten Inhalt (21) des Audio-Puffers (6) sowie die sich daran anschließende Liveübertragung (22) in Text (13) umwandelt und diesen Text (13) mindestens einem Dialogsystem-Prozess (9) zuführt, welcher ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes (13) darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei, falls dies bejaht wird, der Dialogsystem-Prozess (9) eine passende Aktion auslöst oder eine passende Antwort (14) generiert und mit dem Benutzer per Ausgabevorrichtung (3, 4) in Kontakt tritt und

e) dass anderenfalls, falls in den Audio-Daten (11) nur Störgeräusche vorliegen oder falls der Text (13) keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre Spracherkennungs-Prozess (8) und der Dialogsystem-Prozess (9) umgehend wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess (7) zurückgeben.

12. System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus, mit mindestens einem Mikrofon (2), mindestens einem Audio-Puffer (6), mindestens einer Ausgabevorrichtung (3, 4) und einer Hardware-Infrastruktur (25, 26, 27, 28, 29), welche Prozesse (7, 8, 9) ausführen kann, **dadurch gekennzeichnet**, dass die Hardware-Infrastruktur (25, 26, 27, 28, 29) so konfiguriert ist,

a) dass Audio-Daten (11) mit dem mindestens einen Mikrofon (2) aufgenommen werden,

b) dass die Audio-Daten (11) kontinuierlich in dem Audio-Puffer (6) zwischengespeichert werden, so dass der Audio-Puffer (6) stets die Audio-Daten (11) der jüngsten Vergangenheit enthält,

c) dass die Audio-Daten (11) zeitnahe mindestens einem sekundären Spracherkennungs-Prozess (7) zugeführt werden, welcher beim Erkennen eines Schlüsselworts (18) oder einer Phrase mindestens einen primären Spracherkennungs-Prozess (8) startet oder aus einem Ruhezustand aktiviert,

d) dass der primäre Spracherkennungs-Prozess (8) den gesamten oder den jüngsten Inhalt (21) des Audio-Puffers (6) sowie die sich daran anschließende Liveübertragung (22) in Text (13) umwandelt und diesen Text (13) mindestens einem Dialogsystem-Prozess (9) zuführt, welcher ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes (13) darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei, falls dies bejaht wird, der Dialogsystem-Prozess (9) eine passende Aktion auslöst oder eine passende Antwort (14) generiert und mit dem Benutzer per Ausgabevorrichtung (3, 4) in Kontakt tritt und

e) dass anderenfalls, falls in den Audio-Daten (11) nur Störgeräusche vorliegen oder falls der Text (13) keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre Spracherkennungs-Prozess (8) und der Dialogsystem-Prozess (9) umgehend wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess (7) zurückgeben.

Die Patentansprüche 2 bis 11 sind unmittelbar oder mittelbar auf Patentanspruch 1 und die Patentansprüche 13 bis 15 sind unmittelbar oder mittelbar auf Patentanspruch 12 rückbezogen; wegen ihres Wortlauts wird auf die Akte beziehungsweise die Streitpatentschrift verwiesen.

Die Klägerin ist der Ansicht, der jeweilige Gegenstand gemäß den Patentansprüchen 1 und 12 sei gegenüber dem Stand der Technik nicht neu und beruhe nicht auf einer erfinderischen Tätigkeit, was auch für die abhängigen Ansprüche gelte.

Die Klägerin stützt ihr Vorbringen auf folgende Entgegenhaltungen:

**D1** WO 2012/025784 A1,

**D2** US 2005/0216271 A1,

- D3** US 7,424,431 B2,
- D4** US 2012/0010890 A1,
- D5** EP 1 091 346 B1,
- D5a** DE 600 10 827 T2, deutsche Übersetzung der D5,
- D6** US 7,610,199 B2,
- D7** US 6,408,272 B1,
- D8** US 2009/0204410 A1,
- D9** EP 1 091 347 A2.

Die Klägerin beantragt,

das deutsche Patent 10 2013 001 219 für nichtig zu erklären.

Die Beklagte beantragt,

die Klage abzuweisen,

hilfsweise, die Klage abzuweisen,

soweit sie sich auch gegen eine der Fassungen des Streitpatents nach den Hilfsanträgen 1, 1a und 2 bis 5, überreicht mit Widerspruchsbegründung vom 20. Januar 2023 (Hilfsanträge 1 bis 5) bzw. mit Duplik vom 30. März 2023 (Hilfsantrag 1a), richtet,

mit der Maßgabe, dass die Hilfsanträge in der zuvor genannten numerischen Reihenfolge und alle als geschlossene Anspruchssätze gestellt werden.

Die Beklagte tritt der Argumentation der Klägerin entgegen und ist der Auffassung, das Streitpatent sei weder neuheitsschädlich getroffen, noch fehle ihm die erfinderische Tätigkeit. Wenigstens in einer der verteidigten Fassungen nach den eingereichten Hilfsanträgen sei der Gegenstand des Streitpatents schutzfähig.

Wegen des Wortlauts der Ansprüche nach den Hilfsanträgen wird auf die Akte verwiesen.

Die Klägerin tritt auch den Hilfsanträgen entgegen und sieht auch den Gegenstand des Streitpatents in der Fassung der jeweiligen Hilfsanträge als nicht neu und nicht erfinderisch an. Zudem gingen die Hilfsanträge 1 und 4 über die Ursprungsoffenbarung hinaus und Hilfsantrag 4 sei darüber hinaus nicht ausreichend offenbart.

Der Senat hat den Parteien einen Hinweis vom 12. Juni 2023 zugeleitet und hierin Fristen zur Stellungnahme gesetzt.

Wegen der weiteren Einzelheiten des Sach- und Streitstands wird auf die zwischen den Parteien gewechselten Schriftsätze nebst Anlagen, das Protokoll der mündlichen Verhandlung vom 6. März 2024 sowie den weiteren Akteninhalt Bezug genommen.

## **Entscheidungsgründe**

### **A.**

Die zulässige Klage erweist sich als unbegründet. Der Gegenstand des Streitpatents ist in der erteilten Fassung sowohl neu als auch auf erfinderischer Tätigkeit beruhend (§ 22 Abs. 1 PatG i. V. m. §§ 21 Abs. 1 Nr. 1 PatG, 1 bis 5 PatG)

### **I. Zum Gegenstand des Streitpatents in erteilter Fassung, zur Aufgabe, zum Fachmann und zur Auslegung**

1. Der **Gegenstand des Streitpatents** liegt auf dem Gebiet der Spracherkennung, insbesondere der Aktivierung von Vorgängen per Sprache (Streitpatentschrift, Absatz 0001) und betrifft ausweislich der Bezeichnung ein Verfahren und ein System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus.

Smartphones (Mobiltelefone mit Computer-Funktionalität) haben lt. Streitpatent aufgrund ihrer geringen Größe eine stark eingeschränkte Ergonomie, wenn sie per Touchscreen bedient würden. Eine Alternative seien digitale Sprachassistenten, bei denen das Smartphone mit Sprachkommandos, zum Teil auch mit natürlicher Sprache ohne spezielle Steuerbefehle, gesteuert werden könne. Ein bekanntes Beispiel sei das System „Siri“ auf dem Smartphone „iPhone“ des Herstellers Apple (Absatz 0003).

Ein Sprachassistent könne als eine eigenständige App auf dem Smartphone oder in das Betriebssystem integriert sein, so dass die Spracherkennung, Auswertung und Reaktion lokal auf der Hardware des Smartphones erfolgen könnten. In der Regel würde aber wegen der größeren Rechenleistung ein Server-Verbund im Internet („*in the Cloud*“) verwendet, mit dem der digitale Sprachassistent kommuniziere, indem komprimierte Sprach- bzw. Tonaufnahmen an den Server bzw. Server-Verbund geschickt würden und die per Sprachsynthese generierte verbale Antwort zurück auf das Smartphone gestreamt werde (Absatz 0004).

Digitale Sprachassistent-Systeme seien eine Teilmenge der Software-Agenten mit verschiedenen Interaktionsmöglichkeiten, wie z. B. das Abfragen von Fakten, Wissen, Status-Updates in sozialen Netzwerken oder das Diktieren von E-Mails. In den meisten Fällen komme dabei auf der Seite des digitalen Sprachassistenten ein Dialogsystem (bzw. ein sogenannter Chatbot) zum Einsatz, welches zum Teil mit semantischer Analyse oder mit Ansätzen von künstlicher Intelligenz ein realitätsnahes Gespräch zu einem Thema simuliere (Absatz 0005).

Ein konkretes Beispiel für einen digitalen Sprachassistenten sei das als „S Voice“ bezeichnete System auf dem Smartphone „Galaxy S III“ des Herstellers Samsung, bei dem der Benutzer in den Systemeinstellungen bzw. der Systemsteuerung eine begrenzte Anzahl von Phrasen (z. B.: „Hi Galaxy“) hinterlegen könne, die zum Aufwecken aus einem Standby- bzw. Schlafzustand per Sprachbefehl dienen, ohne einen Touchscreen zu berühren oder eine Taste zu drücken. Es sei jedoch nicht möglich, dem Smartphone komplexe Fragen zu stellen oder es zu komplexen Aktionen aufzufordern, solange sich das System im Standby-Modus befinde. Dazu müsse das

Smartphone zuvor explizit per Schlüsselwort aufgeweckt werden und könne erst dann über weitere Kommandos gesteuert werden (Absätze 0006 bis 0008).

2. Dem Streitpatent liegt daher die **Aufgabe** zugrunde, ein Verfahren zu schaffen, mit dem es möglich ist, einem Software-Agenten oder einem digitalen Sprachassistenten, der sich in einem Standby- bzw. Schlafzustand befindet, per „natürlicher“ Sprache komplexe Fragen zu stellen oder auch Mitteilungen und Aufforderungen, wobei das System umgehend und ohne weitere zwischengeschaltete Interaktionsschritte mit einer finalen und vollständigen Antwort oder mit einer Aktion antworten bzw. reagieren soll. Die Komplexität der unterstützten Fragen, Mitteilungen und Aufforderungen soll dabei vergleichbar oder identisch sein mit der Komplexität, die das System im normalen Betrieb beherrscht. Des Weiteren soll das Verfahren durch seine Konzeption besonders vorteilhaft für einen tagelangen Standby-Betrieb des Software-Agenten sein. Für den Benutzer soll der Unterschied zwischen Standby- bzw. Schlafzustand und dem regulären Betrieb kaum bemerkbar sein, d. h. der Benutzer soll den Eindruck gewinnen, dass das System auch im Standby-Modus mit derselben Aufmerksamkeit zuhört wie im regulären Betrieb (Absatz 0009).

Diese Aufgabe soll durch ein Verfahren und ein System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus gemäß den von der Klägerin angegriffenen unabhängigen erteilten Patentansprüchen 1 und 12 gelöst werden.

3. Maßgeblicher **Fachmann** für die Lösung der genannten Aufgabe ist ein Informatiker mit Hochschulabschluss mit mehrjähriger praktischer Berufserfahrung in der Programmierung und Implementierung von Sprachassistenzsystemen und Softwareagenten.

4. Der erteilte Verfahrensanspruch 1 und der unabhängige Vorrichtungsanspruch 12 lauten mit einer Merkmalsgliederung:



## **Anspruch 1**

- 1.1 Verfahren zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus,  
dadurch gekennzeichnet,
- 1.2 a) dass Audio-Daten (11) mit mindestens einem Mikrofon (2) aufgenommen werden,
- 1.3 b) dass die Audio-Daten (11) kontinuierlich in mindestens einem Audio-Puffer (6) zwischengespeichert werden, so dass der Audio-Puffer (6) stets die Audio-Daten (11) der jüngsten Vergangenheit enthält,
- 1.4 c) dass die Audio-Daten (11) zeitnahe [sic!] mindestens einem sekundären Spracherkennungs-Prozess (7) zugeführt werden, welcher beim Erkennen eines Schlüsselworts (18) oder einer Phrase mindestens einen primären Spracherkennungs-Prozess (8) startet oder aus einem Ruhezustand aktiviert,
- 1.5 d) dass der primäre Spracherkennungs-Prozess (8) den gesamten oder den jüngsten Inhalt (21) des Audio-Puffers (6) sowie die sich daran anschließende Liveübertragung (22) in Text (13) umwandelt und diesen Text (13) mindestens einem Dialogsystem-Prozess (9) zuführt,
- 1.6 welcher ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes (13) darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei,
  - 1.6.1 falls dies bejaht wird, der Dialogsystem-Prozess (9) eine passende Aktion auslöst oder eine passende Antwort (14) generiert und mit dem Benutzer per Ausgabevorrichtung (3, 4) in Kontakt tritt und
  - 1.6.2 e) dass anderenfalls, falls in den Audio-Daten (11) nur Störgeräusche vorliegen oder falls der Text (13) keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre

Spracherkennungs-Prozess (8) und der Dialogsystem-Prozess (9) umgehend wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess (7) zurückgeben.

## **Anspruch 12**

- 12.1 System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus, mit
  - 12.1.1 mindestens einem Mikrofon (2),
  - 12.1.2 mindestens einem Audio-Puffer (6),
  - 12.1.3 mindestens einer Ausgabevorrichtung (3, 4) und
  - 12.1.4 einer Hardware-Infrastruktur (25, 26, 27, 28, 29), welche Prozesse (7, 8, 9) ausführen kann,  
*dadurch gekennzeichnet,*
  - 12.1.5 dass die Hardware-Infrastruktur (25, 26, 27, 28, 29) so konfiguriert ist,
- 12.2 a) dass Audio-Daten (11) mit dem mindestens einen Mikrofon (2) aufgenommen werden,
- 12.3 b) dass die Audio-Daten (11) kontinuierlich in dem Audio-Puffer (6) zwischengespeichert werden, so dass der Audio-Puffer (6) stets die Audio-Daten (11) der jüngsten Vergangenheit enthält,
- 12.4 c) dass die Audio-Daten (11) zeitnahe [sic!] mindestens einem sekundären Spracherkennungs-Prozess (7) zugeführt werden, welcher beim Erkennen eines Schlüsselworts (18) oder einer Phrase mindestens einen primären Spracherkennungs-Prozess (8) startet oder aus einem Ruhezustand aktiviert,
- 12.5 d) dass der primäre Spracherkennungs-Prozess (8) den gesamten oder den jüngsten Inhalt (21) des Audio-Puffers (6) sowie die sich daran anschließende Liveübertragung (22) in Text (13) umwandelt

und diesen Text (13) mindestens einem Dialogsystem-Prozess (9) zuführt,

12.6 welcher ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes (13) darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei,

12.6.1 falls dies bejaht wird, der Dialogsystem-Prozess (9) eine passende Aktion auslöst oder eine passende Antwort (14) generiert und mit dem Benutzer per Ausgabevorrichtung (3, 4) in Kontakt tritt und

12.6.2 e) dass anderenfalls, falls in den Audio-Daten (11) nur Störgeräusche vorliegen oder falls der Text (13) keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre Spracherkennungs-Prozess (8) und der Dialogsystem-Prozess (9) umgehend wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess (7) zurückgeben.

5. Der vorstehend definierte zuständige Fachmann versteht die inhaltlich korrespondierenden Gegenstände der im Wesentlichen gleichlautenden, erteilten unabhängigen Ansprüche 1 und 12 und deren Merkmale wie folgt:

#### **5.1 Sprachaktivierung, Software-Agent bzw. Sprachassistent, Standby-Modus (Merkmale 1.1 und 12.1)**

Gemäß den Merkmalen 1.1 und 12.1 der beiden unabhängigen Patentansprüche 1 und 12 nach Hauptantrag wird ein Verfahren bzw. ein System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus beansprucht.

Das Streitpatent geht davon aus, dass Verfahren und Systeme die es ermöglichen, Geräte **per Sprache** zu steuern und auch zu **aktivieren** aus dem Stand der Technik prinzipiell bekannt sind (Absatz 0002).

Als **Software-Agent** wird allgemein eine Software-Einheit bezeichnet, meist ein Computerprogramm, die zu einem gewissen, spezifizierten eigenständigen und eigendynamischen (autonomen) zielgerichteten Verhalten fähig ist, so dass abhängig von verschiedenen Zuständen (Status) ein bestimmter Verarbeitungsvorgang abläuft, ohne dass weitere externe Signale erforderlich sind oder während des Vorgangs ein äußerer Steuerungseingriff erfolgt und der Software-Agent dabei mit seiner Umgebung und ggf. anderen Agenten interagiert. Teile des Software-Agenten können auch als FPGA (Field Programmable Gate Array) oder ASIC (Application-Specific Integrated Circuit) realisiert sein (Absätze 0022 und 0037).

Laut Streitpatent (Absatz 0005) sind als Teilmenge der Software-Agenten digitale **Sprachassistenten** (Sprachassistent-Systeme) zu sehen, wobei das Streitpatent seine technische Lehre in gleicher Weise sowohl für Software-Agenten als auch für digitale Sprachassistenten beschreibt (Absätze 0009, 0011, 0014 und 0015).

Das Streitpatent versteht digitale Sprachassistenten als Alternative zur Bedienung von Smartphones per Touchscreen, die aufgrund ihrer geringen Größe eine stark eingeschränkte Ergonomie hätten (Absatz 0003).

Ein Sprachassistent könne als eigenständige App auf dem Smartphone oder in das Betriebssystem integriert sein, wobei die Spracherkennung, Auswertung und Reaktion lokal auf der Hardware des Smartphones erfolgen könne (Absatz 0004 sowie Figuren 4 und 5 i. V. m. Absätzen 0037 und 0038), aber in der Regel wegen der größeren Rechenleistung ein Server-Verbund im Internet („*in the Cloud*“) verwendet würde, mit dem der digitale Sprachassistent kommuniziere (Absatz 0004 sowie Figur 6 i. V. m. den Absätzen 0039 und 0040).

Das Streitpatent nennt als Beispiele für verschiedene Interaktionsmöglichkeiten mit digitalen Sprachassistenten das Abfragen von Fakten, Wissen, Status-Updates in sozialen Netzwerken oder das Diktieren von E-Mails. In den meisten Fällen käme dabei ein Dialogsystem (bzw. ein sogenannter Chatbot) zum Einsatz, welches zum Teil mit semantischer Analyse oder mit Ansätzen von künstlicher Intelligenz ein realitätsnahes Gespräch zu einem bestimmten Thema simuliere (Absatz 0005).

Unter **Standby-Modus** (auch Ruhe- oder Schlafzustand) versteht der Fachmann den Bereitschaftsbetrieb eines technischen Gerätes, in dem die eigentliche Nutzfunktion temporär deaktiviert ist, aber jederzeit und ohne Vorbereitungen oder längere Wartezeiten wieder aktiviert werden kann, um in der nutzungsfreien Zeit den Energieverbrauch des Gerätes zu senken, da gegenüber dem Normalbetrieb nur eine reduzierte Leistungsaufnahme stattfindet.

Das Streitpatent nennt als Beispiel für einen digitalen Sprachassistenten das als „S Voice“ bezeichnete System auf dem Smartphone „Galaxy S III“ des Herstellers Samsung, bei dem der Benutzer eine begrenzte Anzahl von Phrasen (z. B.: „Hi Galaxy“) hinterlegen könne, die zum Aufwecken aus einem Standby- bzw. Schlafzustand per Sprachbefehl dienen, ohne einen Touchscreen zu berühren oder eine Taste zu drücken (Absätze 0006 bis 0008).

## **5.2 Audio-Puffer; kontinuierlich ... zwischengespeichert; stets ... jüngste Vergangenheit (Merkmal 1.3), zeitnah (Merkmal 1.4); umgehend (Merkmal 1.6.2)**

Gemäß Merkmal 1.3 werden die Audio-Daten kontinuierlich in mindestens einem **Audio-Puffer zwischengespeichert**, so dass der Audio-Puffer **stets** die Audio-Daten der **jüngsten Vergangenheit** enthält und sie nach Merkmal 1.4 **zeitnah** mindestens einem sekundären Spracherkennungs-Prozess zugeführt werden können.

Unter einem **Audio-Puffer** versteht der Fachmann ohne Weiteres eine flüchtige oder nichtflüchtige Speichereinheit, in der die Audiodaten temporär bis zur weiteren Verwendung zwischengespeichert werden.

Ebenso versteht der Fachmann die Angabe „**stets**“ ohne Weiteres in seiner üblichen Bedeutung von „immer“, „kontinuierlich“ oder „jederzeit“.

Dagegen sind die unbestimmten, den zeitlichen Ablauf beschreibenden Begriffe „**jüngsten Vergangenheit**“ und „**zeitnahe**“ bzw. „**ohne nennenswerte Verzögerung**“ (Absatz 0012) sowie „**umgehend**“ vor dem Hintergrund der technischen Lehre des Streitpatents auszulegen, wonach nicht nur Anfragen verarbeitet werden

sollen, die nach Aktivierung des Spracherkennungssystems formuliert werden, sondern in gleicher Weise komplexe Anfragen eines Nutzers, die schon vorher im Standby-Modus des Systems und vor dem Erkennen eines Schlüsselworts geäußert wurden.

Da dabei die in Text umgewandelten Audiosignale daraufhin analysiert werden sollen, ob sie eine Frage, eine Mitteilung und/oder eine Aufforderung enthalten, indem bei einem erkannten Schlüsselwort bzw. einer Phrase, zeitlich rückwärts durchsucht wird, bis ein Zeitabschnitt gefunden ist, welcher sich als Stille bzw. Sprechpause und somit als Anfang eines Satzes interpretieren lässt, liegt die relevante Zeitdauer der „**jüngsten Vergangenheit**“ in der Größenordnung eines typischen gesprochenen Satzes. Das Streitpatent liefert dazu als Beispiele Zeitspannen von mindestens einer Sekunde (Absätze 0030 und 0031), 30 Sekunden (Absätze 0011 und 0026) bis hin zu praxistauglichen Werten zwischen 15 Sekunden und 2 Minuten (Absatz 0024).

Der Fachmann erkennt in diesem Zusammenhang auch, dass es sinnvoll ist, den Dialog des Benutzers mit dem Software-Agenten bzw. Sprachassistenten in Echtzeit über „Live-Audio“-Daten möglichst schnell nach Äußerung eines Schlüsselworts oder einer Phrase aufzunehmen, wofür eine erste Voraussetzung ist, dass die Audio-Daten dem sekundären Spracherkennungs-Prozess zur Erkennung zur Verfügung stehen. Daher wird der Fachmann „**zeitnahe**“ in den Merkmalen 1.4 und 12.4, genauso wie die Angabe „**ohne nennenswerte Verzögerung**“ (Absatz 0012) als nahezu instantan und lediglich durch Verarbeitungsdauern (z. B. durch den A/D-Wandler) des Systems begrenzt verstehen.

Ebenso ist die unbestimmte Zeitangabe „**umgehend**“ zu verstehen, nach welcher der primäre Spracherkennungs-Prozess und der Dialogsystem-Prozess – falls in den Audio-Daten nur Störgeräusche vorliegen oder falls der Text keinen relevanten oder keinen auswertbaren Inhalt enthält – wieder in den Ruhezustand zurückkehren (Absätze 0016, 0036), nämlich als möglichst schnell, d. h. wiederum lediglich durch Verarbeitungsdauern und Übertragungsverzögerungen des Systems begrenzt. In diesem Fall ist der Grund ein möglichst geringer Ressourcen-Verbrauch, den der im

Standby-Modus arbeitende energiesparende sekundäre Spracherkennungs-Prozess gegenüber dem energieintensiveren primären Spracherkennungs-Prozess ermöglicht (Absätze 0012, 0013, 0020, 0021, 0037 und 0038).

Eine zusätzliche Übertragungsverzögerung ergibt sich, wenn – wie als besonders vorteilhaft dargestellt – der primäre Spracherkennungs-Prozess und der Dialogsystem-Prozess auf einem externen Server ausgeführt werden. Diese Latenz soll jedoch durch einen sog. „vorausseilenden Standby-Modus“ reduziert oder nahezu vermieden werden (Figur 6 und Absätze 0039 und 0040).

### **5.3 primärer und sekundärer Spracherkennungs-Prozess (Merkmale 1.4 und 1.5)**

Gemäß Merkmal 1.4 werden die Audio-Daten aus dem Audio-Puffer zeitnah (siehe obigen Abschnitt 5.2) mindestens einem **sekundären Spracherkennungs-Prozess** zugeführt.

Dieser Prozess findet im Standby-Modus statt und startet – bzw. aktiviert aus einem Ruhezustand – einen **primären Spracherkennungs-Prozess** erst, wenn ein Schlüsselwort oder eine entsprechende Phrase in den Audio-Daten erkannt wird. Der primäre Spracherkennungs-Prozess wandelt dann den gesamten oder alternativ nur den jüngsten Inhalt des Audio-Puffers sowie zusätzlich die sich daran anschließende Liveübertragung in Text um (Merkmal 1.5).

Die technische Lehre des Streitpatents beruht im Wesentlichen darauf, dass dem üblichen, im Betriebszustand („Vollbetrieb“) arbeitenden primären Spracherkennungsprozess, der mit einem Dialogsystem-Prozess verknüpft ist, der dauerhaft aktive sekundäre Spracherkennungs-Prozess vorgelagert ist, um die Schlüsselworterkennung auch im Standby- bzw. Ruhezustand des Systems zu gewährleisten.

Der Fachmann entnimmt der Beschreibung die die erteilten Ansprüche nicht beschränkenden Angaben, dass der **sekundäre Spracherkennungs-Prozess** energiesparend, auf geringen Ressourcen-Verbrauch optimiert und als FPGA (Field Programmable Gate Array) oder ASIC (Application-Specific Integrated Circuit) realisiert ist (Absätze 0012, 0021, 0022 und 0037) und dafür nur über einen stark begrenzten

Wortschatz verfügt (Absätze 0023 und 0037) sowie eine erhöhte Falsch-Positiv Rate beim Erkennen von Schlüsselwörtern bzw. Phrasen aufweist (Absätze 0027 und 0036), wobei diese Fehler des sekundären Spracherkennungs-Prozesses von der „hochwertigen“ Spracherkennung des **primären Spracherkennungs-Prozesses** korrigiert werden (Absätze 0020, 0027 und 0036).

Aufgrund des unterschiedlichen Ressourcen-Verbrauchs wird es zudem im Streitpatent als vorteilhaft erachtet, wenn, wie in einer bevorzugten Ausführungsform realisiert, nur der sekundäre Spracherkennungs-Prozess lokal im Endgerät abläuft und der primäre Spracherkennungs-Prozess sowie der Dialogsystem-Prozess auf dem Prozessor eines externen Servers ausgeführt werden, der mit dem Endgerät über ein Netzwerk verbunden ist (Figur 6 i. V. m. Absatz 0017, vorletzter Teilabsatz und Absatz 0039).

#### **5.4 Schlüsselwort, Phrase (Merkmal 1.4)**

Den Begriffen **Schlüsselworte** oder **Phrasen**, welche gemäß Merkmal 1.4 nach Erkennen durch den sekundären Spracherkennungs-Prozess das Starten mindestens eines primären Spracherkennungs-Prozesses auslösen, wird der Fachmann anspruchsgemäß die in der Linguistik üblichen Bedeutungen zuordnen, nämlich Phrase als eine geeignete abgeschlossene syntaktische Einheit verstehen, die aus mehreren Worten besteht, wohingegen es sich bei einem Schlüsselwort nur um ein beliebiges einzelnes Wort handelt, das über seine Funktion als Auslöser als *Schlüsselwort* definiert ist. Ihm ist zudem klar, dass die Schlüsselworte und Phrasen vor der Verwendung des Systems bzw. Ausführung des Verfahrens festgelegt und im System hinterlegt werden müssen.

In der Beschreibung des Streitpatents werden die **Phrasen** in Absatz 0023 so umschrieben, dass sie zugunsten eines geringen Ressourcen-Verbrauchs des sekundären Spracherkennungs-Prozesses aus nur wenigen Wörtern oder kurzen Ausschnitten aus Redewendungen bestehen, die typische Merkmale bei einer Kontaktaufnahme oder einer Frage an den digitalen Sprachassistenten enthalten sollen und sich nicht notwendigerweise am Anfang eines Satzes befinden müssen. Ansonsten werden **Schlüsselwort** und **Phrase** durchgehend als gleichwertig behandelt und



nur durch Beispiele erläutert, insbesondere durch die in den Absätzen 0023 und 0043 exemplarisch genannten Einträge in einem Schlüsselwort- und Phrasen-Katalog:

- Fragewörter und fragende Phrasen: „wer hat“, „was ist“, „wie kann“, „wie ist“, „wo gibt es“, „gibt es“, „weißt du ob“, „kann man“.
- Aufforderungen und Befehle: „Bitte schreibe eine E-Mail an Hans“ oder „Ich möchte ein Foto machen.“, bei denen die Phrase „schreibe eine E-Mail“ bzw. „Foto machen“ erkannt wird.
- Substantive zu Themen, zu denen es Informationen in der Datenbank des Dialogsystems gibt: „Wetter“, „Termin“ und „Fußball“.
- Produktnamen, Spitznamen und Gattungsbegriffe zur direkten Ansprache des digitalen Sprachassistenten: „Handy“, „Smartphone“, „Computer“, „Navi“.

## **5.5 Dialogsystem-Prozess (Merkmal 1.5 und Merkmalsgruppe 1.6)**

Gemäß Merkmal 1.5 des Streitpatents führt der primäre Spracherkennungs-Prozess den in Text umgewandelten gesamten oder jüngsten Inhalt des Audio-Puffers sowie die sich daran anschließende Liveübertragung mindestens einem **Dialogsystem-Prozess** zu.

Unter einem **Dialogsystem-Prozess** versteht der Fachmann allgemein eine Programminstanz, welche eine auf Sprache basierende Kommunikation zwischen dem Systembenutzer und dem System ermöglicht.

Gemäß Merkmal 1.6 analysiert der Dialogsystem-Prozess den Inhalt des Textes daraufhin, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde und gegebenenfalls eine passende Aktion auslöst oder eine passende Antwort generiert und mit dem Benutzer per Ausgabevorrichtung in Kontakt tritt (Merkmal 1.6.1), beispielsweise über Lautsprecher und/oder Display (Absatz 0015), und anderenfalls – falls in den Audio-Daten nur Störgeräusche vorliegen oder falls der Text keinen relevanten oder keinen auswertbaren Inhalt enthält – der Dialogsystem-Prozess zusammen mit dem

primären Spracherkennungs-Prozess sich umgehend (vgl. Abschnitt 5.2) wieder beendet und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess zurückgibt (Merkmal 1.6.2).

Das Streitpatent nennt als Beispiel für ein Dialogsystem einen sogenannten Chatbot, welcher zum Teil mit semantischer Analyse oder mit Ansätzen von künstlicher Intelligenz ein realitätsnahes Gespräch zu einem Thema simuliert (Absätze 0005, 0013, 0014 und 0027).

## **5.6 Text, Inhalt des Textes (Merkmal 1.5 und Merkmalsgruppe 1.6)**

Gemäß Merkmal 1.5 wird der Inhalt des Audio-Puffers sowie die sich daran anschließende Liveübertragung vom primären Spracherkennungs-Prozess in **Text** umgewandelt und der Text mindestens einem Dialogsystem-Prozess zugeführt, welcher diesen hinsichtlich seines **Inhalts** analysiert (Merkmal 1.6; Absätze 0014 und 0033).

Das Streitpatent gibt in der Beschreibungseinleitung das Umwandeln von akustischen Sprachsignalen in Text, konkret, das Umwandeln in eine digitale Text-Darstellung mittels einer Zeichenkodierung, als prinzipiell bekannt an (Absatz 0002).

Ansonsten wird im Streitpatent, ebenfalls die Ansprüche nicht beschränkend, lediglich erläutert, dass Text per Spracherkennung erzeugt wird (Absatz 0013) und als Text-String (Absatz 0016), d. h. als Zeichenkette einer endlichen Folge von Zeichen bzw. Symbolen (z. B. Buchstaben, Ziffern, Sonderzeichen und Steuerzeichen) aus einem definierten Zeichensatz vorliegt sowie im Dialogsystem beispielsweise mittels semantischer Analyse oder mit Ansätzen von künstlicher Intelligenz analysiert werden kann (Absatz 0027).

Der Fachmann wird daher die Anforderungen an das Format des zu analysierenden Textes derart funktionell verstehen, dass es geeignet sein muss die Audio-Daten dahingehend zu analysieren, ob sie bestimmte sprachliche Einheiten des Benutzers enthalten, die einer Frage, einer Mitteilung und/oder einer Aufforderung entsprechen, was einen Vergleich mit den hinterlegten Inhalten voraussetzt. Dem Fachmann sind dazu Techniken der automatischen Spracherkennung wie beispielsweise

Fast-Fourier-Transformation, Hidden-Markov-Modelle und künstliche neuronale Netzwerke geläufig (vgl. beispielsweise Druckschrift D1, Seite 1 und Seite 17, jeweils letzter Absatz sowie Druckschrift D2, Absätze 0024 und 0048).

## II. Zur Patentfähigkeit

Dem Streitpatent in der erteilten Fassung steht der Nichtigkeitsgrund der fehlenden Patentfähigkeit nach § 22 Abs. 1 PatG i. V. m. § 21 Abs. 1 Nr. 1 PatG, §§ 1 bis 5 PatG nicht entgegen. Denn die hiermit unter Schutz gestellte Lehre erweist sich gegenüber dem im Verfahren entgegengehaltenen Stand der Technik als neu und auf erfinderischer Tätigkeit beruhend.

1. Das Verfahren des erteilten Patentanspruchs 1 ist **neu** gegenüber dem Stand der Technik nach der Druckschrift **D2 (US 2005/0216271 A1)**, da ihr insbesondere die Merkmale M1.5 und M1.6.2 nicht vollständig entnehmbar sind. Ausgehend von diesem nächstliegenden Stand der Technik ist das Verfahren des Patentanspruchs 1 dem Fachmann weder allein in Verbindung mit seinem Fachwissen noch in der Zusammenschau mit einer oder mehreren der weiteren im Verfahren befindlichen Druckschriften nahegelegt, so dass es diesem gegenüber auch auf einer **erfinderischen Tätigkeit beruht**.

1.1 Die Druckschrift **D2** betrifft laut der Bezeichnung ein Sprachdialogsystem („*speech dialogue system*“, SDS) für die Steuerung eines elektronischen Geräts. Derartige Sprachdialogsysteme seien in der Lage, von einem Benutzer gesprochene Anweisungen in akustischen Signalen zu analysieren, um Steuerbefehle zu bestimmen, die dann von dem elektronischen Gerät ausgeführt werden (Absätze 0004 und 0006).

Sprachgesteuerte Geräte seien besonders nützlich in Umgebungen, in denen die Hände des Benutzers für andere Tätigkeiten, z. B. zum sicheren Fahren eines Fahrzeugs benötigt würden und trotzdem Autoradio, Telefon, Klimaanlage oder Naviga-

tionssystem bedient werden sollen. Da es bei herkömmlichen Sprachdialogsystemen aber erforderlich sein könne, dass der Benutzer vor der Eingabe von Steuerbefehlen über eine Spracheingabe eine spezielle Taste drücken müsse (Push-to-Talk), um einen Sprachdialog mit dem Sprachdialogsystem zu initiieren, bestehe ein Bedarf, ein Sprachdialogsystem zum Steuern eines elektronischen Geräts mit verbesserter Benutzerfreundlichkeit bereitzustellen (Absätze 0007 bis 0009).

Das dazu in der Druckschrift **D2** u. a. vorgeschlagene Verfahren zum Steuern eines elektronischen Geräts mittels eines Sprachdialogsystems (SDS) mit einem Spracherkennungssystem verwendet eine Steuerbefehl-Bestimmungseinheit, die durch ein von einem Benutzer des SDS gesprochenes Schlüsselwort („*keyword*“) aktiviert wird (Absatz 0010).

**1.2** Aus der Druckschrift **D2** ist in Worten des erteilten Anspruchs 1 ausgedrückt Folgendes bekannt: Ein

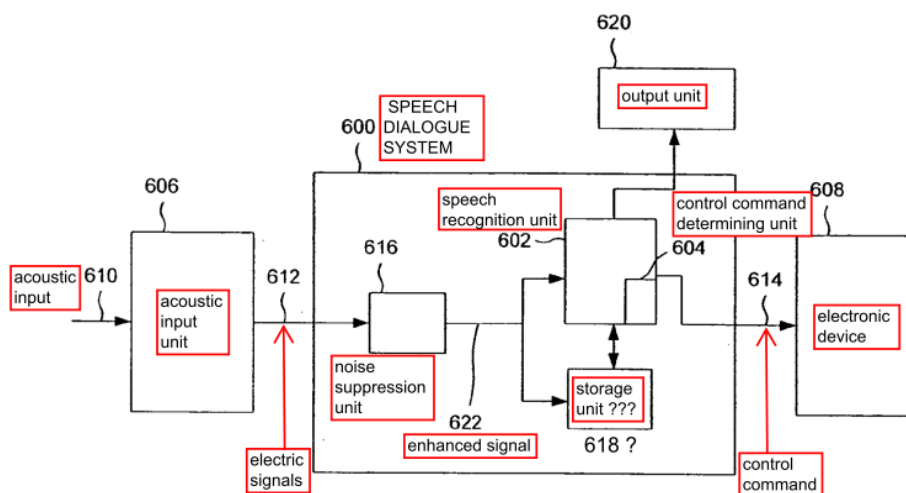
1.1 Verfahren zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus, wobei

Die Druckschrift D2 beschreibt die Aktivierung eines Software-Agenten bzw. Sprachassistenten eines Sprachdialogsystems (SDS) zum Steuern eines elektronischen Geräts (Absatz 0004: „*a speech dialogue system for controlling an electronic device using a speech dialogue system.*“), wie beispielsweise ein Mobiltelefon, ein Audio- und/oder Videounterhaltungssystem, ein Navigationssystem oder eine Klimaanlage in einem Fahrzeug (Absätze 0046 und 0055).

Da die für den Dialog mit dem Benutzer verantwortliche Steuerbefehl-Bestimmungseinheit („*control command determining unit (504 or 604)*“) im Rahmen des Verfahrens erst durch die Eingabe und das Erkennen von Schlüsselwörtern („*keywords*“) aktiviert werden muss, erfolgt die Aktivierung durch Sprache aus dem Ruhezustand bzw. einem Standby-Modus heraus (Absatz 0010: „*The speech recognition system may comprise a control com-*

mand determination unit to be activated by a keyword for determining a control command to control an electronic device.“, Absatz 0027: „Keyword information to activate the control command determining unit 504“, Absatz 0028: „If the speech recognition unit 502 has identified a keyword, the control command determining unit 504 may be activated 104 and start searching for control command information in the part of the input that comes after the keyword information 105.“ und Absatz 0040: „the keyword for activating the control command determining unit (504 or 604)“).

Figur 6 mit der zugehörigen Beschreibung in den Absätzen 0051 bis 0055 zeigt bzw. beschreibt ein Ausführungsbeispiel eines Sprachdialogsystems zur Ausführung eines entsprechenden Verfahrens.



Figur 6 der Druckschrift D2 mit Ergänzung durch den Senat

- 1.2 a) Audio-Daten („acoustic inputs“ 510; 610) mit mindestens einem Mikrofon aufgenommen werden,

Absatz 0045: „FIG. 5 illustrates a speech dialogue system for controlling an electronic device (SDS) 500. The SDS 500 may receive acoustic inputs 510, which may comprise acoustic inputs from a user of the SDS 500 via an acoustic input unit 506. The acoustic input unit may comprise, for example, a microphone.“ und Absatz 0052: „microphones used in the acoustic input unit

*606 may be directional microphones that receive signals from the direction of the positions of the occupants of the car.”.*

- 1.3 b) dass die Audio-Daten (510; 610) kontinuierlich in mindestens einem Audio-Puffer („*storage unit*“ 518; 618) zwischengespeichert werden, so dass der Audio-Puffer stets die Audio-Daten der jüngsten Vergangenheit enthält,

Die Audio-Daten der jüngsten Vergangenheit werden in einer als Audio-Puffer fungierenden temporären Speichereinheit (518; 618) zwischengespeichert (Absatz 0022: „*The received signals may then be stored 101 by the storage unit 618 (See FIG. 6). The storage unit 618 may be a cacheable memory such as volatile memory chips or flash memory units to allow fast access by the SDS 500. By providing the storage unit 518, the acoustic input 510 input during a predetermined time duration may be kept in the memory of the SDS 500 so that analysis is faster as compared to on-the-fly-only analysis.”).*

Die Audio-Daten werden dabei kontinuierlich in diesem als FIFO-Speichereinheit beschriebenen Audio-Puffer zwischengespeichert, so dass er immer die Audio-Daten der jüngsten Vergangenheit enthält (Absatz 0026: „*Alternatively, the SDS 500 may check 310 (see FIG. 3) whether new signals have been stored in the storage unit 618 and determine whether these new signals contain keywords 311.*“ und insbesondere Absatz 0053: „*The storage unit 618 may be configured to store data corresponding to a predetermined time interval and, thus, continuously remove the earliest entry to add new incoming data.*“).

- 1.4 c) dass die Audio-Daten (510; 610) zeitnah mindestens einem sekundären Spracherkennungs-Prozess (502; 602) zugeführt werden, welcher beim Erkennen eines Schlüsselworts („*keyword*“) oder einer Phrase („*a combination of several words*“) mindestens einen

primären Spracherkennungs-Prozess (504; 604) startet oder aus einem Ruhezustand aktiviert,

Die Audio-Daten werden einer Spracherkennungs-Einheit 502 bzw. 602 zugeführt (Figuren 5 und 6 i. V. m. Absatz 0023: „*The speech recognition unit 502 may read the stored input or at least a part of the stored input from the storage unit 618.*“ und Absatz 0053: „*the storage unit 618, from which the speech recognition unit 602 may receive signals.*“), welche einen sekundären Spracherkennungs-Prozess ausführt, indem sie die Audio-Daten analysiert, um Schlüsselworte oder Phrasen zu identifizieren (Absatz 0025: „*The speech recognition unit 502 may identify the received keywords by comparing the words or a combination of the words with a first-vocabulary set comprising keyword vocabulary elements to determine whether the user has pronounced one of the keywords present in the first vocabulary set 103. The SDS 500 may continuously analyze the received acoustic signals to check whether the user has pronounced keyword information, thus indicating to the SDS that he wants to provide a control command for an electronic device 508. The first vocabulary set may comprise keywords, where one keyword may actually be a combination of several words.*“, Absatz 0053: „*the speech recognition unit 602 where keyword information may be searched.*“ und Absatz 0054: „*Once the speech recognition unit 602 has identified a keyword and/or a control command*“).

Beim Erkennen eines Schlüsselworts oder einer Phrase wird die Steuerbefehl-Bestimmungseinheit („*control command determining unit 504 or 604*“) und damit ein *primärer* Spracherkennungs-Prozess aus einem Ruhezustand aktiviert (Absatz 0027: „*Keyword information to activate the control command determining unit 504 may be provided in a number of ways. Several words or a combination of words can be used.*“, Absatz 0028: „*If the speech recognition unit 502 has identified a keyword, the control command determining unit 504 may be activated 104*“, Absatz

0040: „the keyword for activating the control command determining unit (504 or 604)”, Absatz 0047: „*If the input contains keyword information, the control command determining unit 504 may be activated*” und Absatz 0049: „... *if the speech recognition unit 502 recognizes a keyword out of the provided keyword information that is part of the acoustic input 510, the control command determining unit 504 is activated.*”).

1.5<sup>Teil</sup> d) dass der primäre Spracherkennungs-Prozess (504; 604) den gesamten oder den jüngsten Inhalt des Audio-Puffers (518; 618) sowie die sich daran anschließende Liveübertragung ~~in Text umwandelt und diesen Text~~ mindestens einem Dialogsystem-Prozess zuführt,

Die den primären Spracherkennungs-Prozess ausführende Steuerbefehl-Bestimmungseinheit (504; 604) führt den Inhalt des Audio-Puffers einem integrierten Dialogsystem-Prozess zu (Figuren 5 und 6 i. V. m. Absatz 0028: „*If the speech recognition unit 502 has identified a keyword, the control command determining unit 504 may be activated 104 and start searching for control command information in the part of the input that comes after the keyword information 105. Similar to the way the keyword is found, a control command may be identified by comparing determined words or a combination thereof with a predetermined second vocabulary set of control command vocabulary elements to identify a control command out of the provided control command information.*” und den Absätzen 0029 bis 0032). Der Dialogsystem-Prozess („*speech dialogue*“) wird in den Absätzen 0036 bis 0044 an Beispielen erläutert.

Zwar wird dem primären Spracherkennungs-Prozess (504; 604) und damit dem Dialogsystem-Prozess dabei sowohl der gesamte oder der jüngste Inhalt des Audio-Puffers (Absatz 0023: „*The speech recognition unit 502 may read the stored input or at least a part of the stored input from the storage unit 618.*“) als auch



eine sich daran anschließende Liveübertragung (Absatz 0023: „*The speech recognition unit 502 may also directly receive the acoustic input from the acoustic input unit 506 or from the noise suppression unit 616.*“) im Sinne des Streitpatents zugeführt. Dass der primäre Spracherkennungs-Prozess (504; 604) auf den Inhalt des Audio-Puffers zugreifen können muss und diesen dem Dialog-Prozess zuführt, ergibt sich auch aus den Dialog-Beispielen, in denen Befehlskontrollinformation („*control command*“, „*second vocabulary*“) vor dem Schlüsselwort („*keyword*“, „*first-vocabulary*“) in den aufgenommenen Audio-Daten enthalten ist (Absatz 0032: „*the control command determining unit 504 may start searching the acoustic input 510 prior to the keyword information to check whether the user made the control command prior to the keyword 108.*“).

Nicht entnehmbar ist dem Verfahren gemäß Druckschrift D2 jedoch, dass der primäre Spracherkennungs-Prozess (504; 604) den Inhalt des Audio-Puffers (518; 618) sowie die sich daran anschließende Liveübertragung in Text umwandelt. Zwar werden die aufgenommenen Audio-Daten zur Durchführung der Spracherkennung in einen analysierbaren Text im Sinne des Streitpatents (vgl. Auslegung in Abschnitt I. 5.6) umgewandelt (Absatz 0024: „*To carry out the speech recognition, the speech recognition unit 502 may analyze the received electric signal by performing, for example, a Fourier transform. The recognition may be based on a hidden Markov model or neural networks capable of determining words out of a continuous speech input comprising more than one word.*“ und Absatz 0048: „*To perform speech recognition, the speech recognition unit 502 may analyze the received electric signal by performing, for example, a Fourier transform. The recognition may be based on a hidden Markov model or neural networks capable of identifying words out of a continuous speech input comprising more than one word.*“). Diese Um-

wandlung findet jedoch bereits in der dem permanent aktiven sekundären Spracherkennungs-Prozess entsprechenden „*speech recognition unit 502*“ bzw. 602 statt und wird nicht vom erst zu aktivierenden primären Spracherkennungs-Prozess (504; 604) durchgeführt.

Auch die Angaben in den Absätzen 0032, 0033, 0047 und 0049, versteht der Fachmann nicht dahingehend, dass sowohl die übergeordnete (502; 602) als auch die Subkomponente (504; 604) jeweils eine entsprechende Umwandlung in Text vornehmen, wie die Klägerin meint. Insbesondere versteht der Fachmann die Angaben im Absatz 0032 („*the control command determining unit 504 may start searching the acoustic input 510*“), im Absatz 0033 („*the control command determining unit 504 may also be configured to search the acoustic input 510*“), im Absatz 0047 („*the control command determining unit 504 may be activated to look for control command information contained in the acoustic input 510.*“) und im Absatz 0049 („... *the control command determining unit 504 is activated. Then the acoustic input 510 may be searched for control command information.*“) selbstverständlich nicht derart, dass die Steuerbefehlsbestimmungseinheit 504 bzw. 604 die akustische Eingabe 510 bzw. 610 direkt durchsucht. Vielmehr entnimmt er den Figuren 5 und 6 i. V. m. den zugehörigen Beschreibungsteilen der Absätze 0045 und 0052, dass es sich beim „*acoustic input*“ 510 bzw. 610 um die akustische Äußerung eines Benutzers, d. h. ein „Schallsignal“ handelt, noch bevor dieses in einer akustischen Eingabeeinheit 506 bzw. 606 von einem Mikrofon in ein elektrisches Signal umgewandelt wird und die in dieser Form aus physikalischen Gründen prinzipiell nicht von einer der Komponenten des Sprachdialogsystems (SDS) analysiert werden kann. Für den Fachmann ist daher offensichtlich, dass nicht etwa das akustische „Rohsignal“ 510 bzw. 610 durchsucht wird, sondern nur ein geeignetes elektrisches, digitalisiertes und in Text umgewandeltes Signal.

Dies erkennt er vor allem anhand der Figur 6 i. V. m. den Absätzen 0051 bis 0053, die den Signalpfad des „*acoustic input 610*“ zeigen bzw. beschreiben, wo nach Umwandlung des akustischen in ein „*electric signal 612*“ mit dem Mikrofon der „*acoustic input unit 606*“ und nach der Rauschunterdrückung in der „*noise suppression unit 616*“ dieses als „*enhanced signal 622*“ der Spracherkennungseinheit 602 zugeführt wird, entweder direkt (oberer horizontaler Pfeil) oder indirekt über die Speichereinheit 618 (unterer horizontaler Pfeil und vertikaler Doppelpfeil), wohingegen kein Signalpfad das „*enhanced signal 622*“ zur Steuerbefehlsbestimmungseinheit 604 führt.

- 1.6 welcher ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei,

Sobald der primäre Spracherkennungs-Prozess (504; 604) und der Dialogsystem-Prozess aktiviert sind (vgl. Ausführungen zu Merkmal 1.4), werden die vorliegenden Audio-Daten bzw. der Inhalt des Textes darauf hin analysiert, ob Steuerbefehlsinformationen („*control command information*“) enthalten sind, d. h. es werden ähnlich wie bei der Schlüsselwort-Erkennung, bestimmte Wörter oder eine Wortkombination mit einem vorbestimmten zweiten Vokabelsatz abgeglichen, um als Steuerbefehl identifiziert werden zu können (Absatz 0028: „*If ... the control command determining unit 504 may be activated 104 and start searching for control command information in the part of the input that comes after the keyword information 105. Similar to the way the keyword is found, a control command may be identified by comparing determined words or a combination thereof with a predetermined second vocabulary set of control command vocabulary elements to identify a control command out of the provided control command information.*“). Die gesuchten Steuerbefehlsinformationen sind beispielsweise vom Benutzer an den Software-

Agenten gerichtete Mitteilungen (z. B. Absatz 0037: „*SDS, I want to phone.*“) oder Aufforderungen (z. B. Absatz 0033: „*change the radio station, car*“, Absatz 0031: „*show me the way home.*“).

1.6.1 falls dies bejaht wird, der Dialogsystem-Prozess eine passende Aktion auslöst oder eine passende Antwort generiert und mit dem Benutzer per Ausgabevorrichtung in Kontakt tritt und

Falls die Analyse ergibt, dass mit dem gefundenen Steuerbefehl eine Mitteilung und/oder eine Aufforderung vorliegt, löst der Dialogsystem-Prozess eine passende Aktion aus (z. B. Absatz 0038: „*Based on the keyword and control command determination, the SDS will inform the telephone that the user is going to make a phone call and at the same time may ask the user to provide the telephone number he wants to call.*“, Absatz 0040: „*the control command “telephone” and will carry out the same actions as described.*“) und/oder es wird ein Dialog gestartet und per Ausgabevorrichtung (Absatz 0051: „*... the communication with an acoustical or optical output unit 620, such as loudspeakers or a display.*“) eine passende Antwort an den Benutzer generiert (z. B. Absatz 0035: „*output a message 213 via an acoustic or optic output unit 620 ... Typical messages may include “speech dialogue system turned on, do you want to proceed” or “the speech dialogue system determined that you wish to change the radio station to FM 94.8, please confirm.*“).

1.6.2<sup>Teil e</sup>) dass anderenfalls, falls in den Audio-Daten nur Störgeräusche vorliegen oder falls der Text keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre Spracherkennungs-Prozess und der Dialogsystem-Pro-

zess ~~umgehend~~ wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess zurückgeben.

Obwohl es in der Druckschrift D2 nicht explizit beschrieben wird, entnimmt der Fachmann aus der Tatsache, dass der primäre Spracherkennungs-Prozess (504; 604) und der Dialogsystem-Prozess erst aktiviert werden müssen, dass sie nicht permanent in Betrieb sind und daher auch wieder in den Ruhezustand zurückkehren oder sich beenden, falls in den Audio-Daten nur Störgeräusche vorliegen oder falls der Text keinen relevanten oder keinen auswertbaren Inhalt enthält. Daraus kann jedoch nicht zwingend auf eine „umgehende“ Rückkehr in den Ruhezustand im Sinne des Streitpatents geschlossen werden. Vielmehr kann beispielsweise auch, wie beim fachüblichen „Beenden-bei-Zeitüberschreitung“-Verfahren („*Timeout*“), eine bestimmte Zeitdauer auf erneute Audio-Eingaben oder das Eintreten einer anderen Bedingung gewartet werden, insbesondere um nicht unnötig oft Aktivierungs- / Deaktivierungs-Zyklen durchlaufen zu müssen.

Da die Druckschrift D2 somit die Merkmale 1.5 und 1.6.2 zumindest nicht vollständig offenbart, ist das Verfahren des erteilten Patentanspruchs 1 gegenüber dem Stand der Technik nach der Druckschrift D2 **neu**.

**1.3** Das Verfahren gemäß Patentanspruch 1 nach Streitpatent beruht gegenüber dem Stand der Technik nach der Druckschrift **D2** auch auf einer **erfinderischen Tätigkeit**.

Es ist bereits fraglich, ob ein Fachmann, der vor der Aufgabe steht ein Verfahren bereitzustellen, mit dem ein Software-Agent mittels Sprache aus einem Standby-Modus aktiviert werden kann, um so einen energiesparenden tagelangen Standby-

Betrieb zu ermöglichen (vgl. Streitpatentschrift, Absatz 0009), von dem Sprachdialogsystem der Druckschrift D2 ausgehen würde, welches die Problematik des Energieverbrauchs an keiner Stelle thematisiert und daher auch keine Lösung dafür vorschlägt.

Doch selbst wenn der Fachmann D2 als Ausgangspunkt wählen würde, ist auch aufgrund dessen keine Veranlassung für ihn erkennbar, den fehlenden **Teil des Merkmals 1.6.2** zu realisieren und den primären Spracherkennungs- und den Dialogsystem-Prozess der „*control command determining unit 504*“ umgehend wieder in den Ruhezustand zurückkehren zu lassen bzw. zu beenden.

Denn da das Sprachdialogsystem (SDS) gemäß D2 vor allem zum Steuern von elektronischen Geräten wie Autoradio, Telefon, Klimaanlage, DVD-Player, Audiounterhaltungs- oder Navigationssystem in einer Fahrzeugumgebung (Absätze 0007, 0011, 0029, 0031, 0042, 0044 bis 0046, 0052 und 0055) eingesetzt werden soll, spielt die Frage der Energieeffizienz und der sparsame Umgang mit der Akkumulatorladung des Sprachdialogsystems für einen langen Betrieb keine oder nur eine sehr untergeordnete Rolle, da die Energie durch die „Batterie“, mit in der Regel relativ großer Kapazität, verbunden mit der Lichtmaschine des Fahrzeugs, nahezu unbegrenzt zur Verfügung steht. Somit ist es nachvollziehbar, dass in der D2 nicht angegeben wird, ob überhaupt und falls ja, zu welchem Zeitpunkt und durch welchen Auslöser die entsprechenden Prozesse wieder deaktiviert werden sollen.

Der Fachmann der ausgehend von der D2 eine definierte Rückkehr in den Ruhezustand im Sprachdialogsystem implementieren möchte, muss mangels diesbezüglicher Hinweise oder Anregungen daher eigene Überlegungen anstellen und stellt fest, dass neben dem fachüblichen Beenden-bei-Zeitüberschreitung-Verfahren („*Timeout*“) durch Warten auf einen voreingestellten Zeitablauf nicht nur die eine Alternative der *umgehenden* Prozessbeendung existiert, sondern ebenso als weitere Alternative das Warten auf erneute Audio- oder andere Eingaben oder das Eintreten einer anderen Bedingung sowie vor allem – besonders vorteilhaft beim Sprachsteuern von elektronischen Geräten in Fahrzeugen – nach dem Starten aus

dem Standby-Betrieb ein Beenden erst beim Beenden der Fahrt, beispielsweise beim Abstellen des Fahrzeugmotors, die er ebenso in Erwägung zieht.

Daher erweist sich das umgehende Beenden nach dem fehlenden Teil des Merkmals 1.6.2 nicht als naheliegende Maßnahme, zumal bei den genannten Alternativen des nicht umgehenden Beendens unnötige Aktivierungs-/Deaktivierungs-Zyklen und die damit einhergehende erhöhte Fehlerrate vermieden bzw. vermindert werden können.

Weiterhin ist es für den Fachmann nicht naheliegend, ausgehend von der technischen Lehre der Druckschrift D2 den fehlenden **Teil des Merkmals 1.5** zu realisieren und die Audiodaten aus dem Audio-Puffer und der Liveübertragung vom primären Spracherkennungs-Prozess in Text umwandeln zu lassen.

Wie oben in Abschnitt 1.2 zum Merkmal 1.5 dargelegt, findet gemäß D2 die Umwandlung des Inhalts des Audio-Puffers (518 bzw. 618) sowie der sich daran anschließenden Liveübertragung in Text bereits in der den permanent aktiven sekundären Spracherkennungs-Prozess ausführenden Spracherkennungseinheit 502 bzw. 602 statt und wird nicht vom erst zu aktivierenden primären Spracherkennungs-Prozess der Steuerbefehlsbestimmungseinheit 504 bzw. 604 durchgeführt.

Um zur Maßnahme des fehlenden Teils von Merkmal 1.5 zu gelangen bestehen prinzipiell zwei Möglichkeiten:

- a) Die Umwandlung in Text durch den sekundären und zusätzlich durch den primären Spracherkennungs-Prozess oder
- b) durch den primären Spracherkennungs-Prozess statt durch den sekundären Spracherkennungs-Prozess durchzuführen.

Die erste Alternative a) ist weder der Druckschrift D2 entnehmbar – wie die Klägerin vorträgt – noch für den Fachmann naheliegend. Denn es ist keine Veranlassung erkennbar, warum der primäre Spracherkennungs-Prozess in der Steuerbefehlsbestimmungseinheit 504 bzw. 604, die als Teilkomponente der Spracherkennungseinheit 502 bzw. 602 ausgebildet ist (vgl. die schematische Darstellung in Figur 6

i. V. m Absatz 0051: „a speech recognition unit 602, which may further comprise a control command recognition unit 604.“), erneut eine Umwandlung in Text vornehmen sollte, wenn dieser Text bereits in der ihr übergeordneten Einheit vorliegt. Ein solches Vorgehen wäre unökonomisch, denn es würde Rechenleistung im SDS binden, unnötigerweise zusätzliche Energie verbrauchen, ggf. zu längeren Latenzzeiten führen und würde daher vom Fachmann nicht in Erwägung gezogen.

Die zweite Alternative b) ist der D2 ebenfalls nicht zu entnehmen und für den Fachmann ebenso fernliegend, da ein Tausch der Umwandlung von Audio-Daten in Text zwischen den beiden Einheiten bzw. Prozessen zum einen eine Ausstattung der Steuerbefehlsbestimmungseinheit 504 bzw. 604 mit einer geeigneten Hard- und Software erfordern würde und zum anderen, in der Spracherkennungseinheit 502 bzw. 602 eine andere Methode bereitzustellen, mit der weiterhin eine Schlüsselwortidentifikation sichergestellt werden könnte. Einen Hinweis oder eine Anregung für derartige umfassende Änderungen des Gesamtkonzepts kann der Fachmann der Druckschrift D2 nicht entnehmen zumal damit kein erkennbarer Vorteil verbunden wäre.

Da der Fachmann aufgrund seines Fachwissens somit keine Veranlassung hat, die in sich geschlossene technische Lehre der Druckschrift D2 zu verlassen, kann auch eine Berücksichtigung des Weiteren im Verfahren befindlichen Standes der Technik ihn nicht veranlassen, die Maßnahmen der fehlenden Teilmerkmale 1.5 und 1.6.2 zu realisieren.

Der Fachmann kann daher ausgehend von der Druckschrift D2 nicht zum Gegenstand des Patentanspruchs 1 gelangen, ohne erfinderisch tätig zu werden.

**2.** Das Verfahren des erteilten Patentanspruchs 1 ist **neu** gegenüber dem Stand der Technik nach der Druckschrift **D4 (US 2012/0010890 A1)**, da ihr insbesondere die Merkmale M1.3, 1.4 und M1.6.2 nicht vollständig entnehmbar sind. Ausgehend von diesem Stand der Technik ist das Verfahren des Patentanspruchs 1 dem Fach-



mann weder allein in Verbindung mit seinem Fachwissen noch in der Zusammenchau mit einer oder mehreren der weiteren vorliegenden Druckschriften nahegelegt, so dass es auch dieser gegenüber auf einer **erfinderischen Tätigkeit beruht**.

**2.1** Die Druckschrift **D4** beschreibt ein immer eingeschaltetes, sprachaktivierbares, leistungsoptimiertes drahtloses Kommunikationsgerät mit zugehöriger Basis, das eine Person rund um die Uhr freihändig nutzen kann, entweder um den Hals oder am Handgelenk oder wo immer es ihren Bedürfnissen am besten entspricht. Die Sprachaktivierung bietet der Person einen größeren Komfort und ermöglicht dem Mikrocontroller gleichzeitig eine bessere Kontrolle der energieverbrauchenden Ressourcen. Darüber hinaus befinden sich Anwendungen in der Basis und auf einem Anwendungsserver (entweder Sprache oder Daten), auf den die drahtlose Kommunikationsbasis zugreift (Abstract).

Als Alternative zur manuellen und im Allgemeinen komplexen Interaktion über winzige Tastaturen und winzige Displays bei heutigen drahtlosen Geräten mit einer Vielzahl von Funktionen und integrierten Anwendungen, sei die kontinuierliche Spracherkennung auf einem drahtlosen Kommunikationsgerät problematisch, da hierfür viel Rechenleistung des Computers erforderlich sei, was wiederum zu relativ kurzen Nutzungszeiten zwischen dem Aufladen des Akkus führe (Absatz 0014). Somit bestünde das Ziel für ein modernes drahtloses Kommunikationsgerät darin, dem Benutzer einen möglichst breiten Zugriff auf Anwendungen und Informationen zu bieten und gleichzeitig die Notwendigkeit eines häufigen Aufladens des Akkus zu minimieren (Absatz 0015).

**2.2** Die Druckschrift **D4** offenbart dem Fachmann in Worten des erteilten Anspruchs 1 nicht mehr als Folgendes: Ein

1.1 Verfahren zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus, wobei

Abstract: „*an Always On, Hands-free, Speech Activated, Power-optimized Wireless Communications Device*“, Absatz

0038: „*In the idle 303 state, the microcontroller 205 can receive an indication 215 that a keyword phrase was spoken by the user 101 as detected by the speech recognition 209 circuit. The microcontroller 205 will wake up from its sleep and enter the keyword recognized 312 state.*“ sowie Absatz 0025: „*Additionally, the base provides access to a variety of external media servers 108 supporting audio-based applications*“.

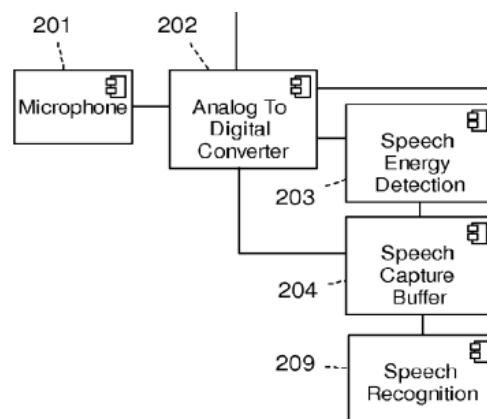
- 1.2 a) dass Audio-Daten mit mindestens einem Mikrofon (201) aufgenommen werden,

Figur 2 i. V. m. Absatz 0026: „*The microphone 201 continuously picks up sound in the immediate environment.*“

- 1.3<sup>Teil</sup> b) dass die Audio-Daten ~~kontinuierlich~~ in mindestens einem Audio-Puffer (204) zwischengespeichert werden, so dass der Audio-Puffer ~~stets~~ die Audio-Daten der jüngsten Vergangenheit enthält,

Die Audio-Daten werden zwar mit dem Mikrofon (201) kontinuierlich aufgenommen und nach der Digitalisierung in einem ADC („*analog to digital converter or ADC 202*“) einer Sprachenergieerkennungsschaltung („*speech energy detection circuit 203*“) zugeleitet. In dem nachfolgenden, einem Audio-Puffer des Streitpatents entsprechenden Spracherfassungspuffer 204 („*speech capture buffer 204*“) werden sie aber nur dann zwischengespeichert, wenn die Sprachenergie über einem vordefinierten Schwellenwert liegt (Absatz 0026: „*The speech energy detection 203 circuit continuously analyzes the ADC 202 output and over a number of successive samples determine if there is speech energy detected above a predefined threshold. The speech capture buffer 204 circuit is engaged at the beginning of detecting sufficient speech energy, and is turned off when detected speech energy drops below the predefined threshold. While engaged, the speech capture buffer 204 stores all output from the ADC 202. Additionally, the speech capture buffer 204 can be continuously*

engaged in recording speech data for use in other applications." i. V. m. Figur 2 sowie Anspruch 1, Gliederungspunkte o) bis q)).



Ausschnitt aus Figur 2 der Druckschrift D4

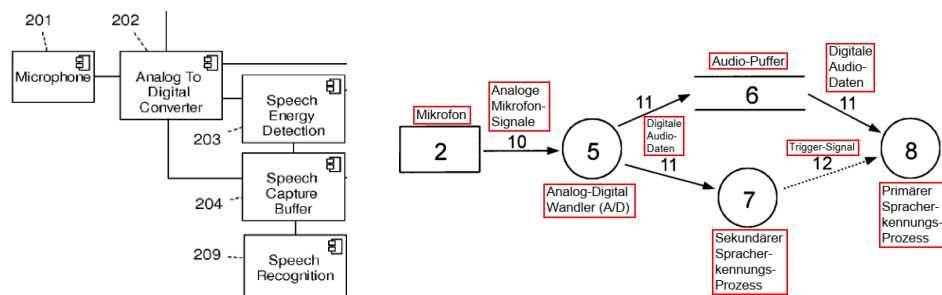
Somit ist das Merkmal 1.3 in der Druckschrift D4 nicht vollständig offenbart, da der Audio-Puffer („*speech capture buffer* 204“) nicht stets die Audio-Daten der jüngsten Vergangenheit enthält, da diese nicht kontinuierlich im Audio-Puffer 204 zwischengespeichert werden, sondern nur dann, wenn die Sprachenergieerkennungsschaltung 203 die Sprachenergie als hoch genug bestimmt und den Audio-Puffer 204 aktiviert.

1.4<sup>Teil</sup> c) dass die Audio-Daten zeitnah mindestens einem sekundären Spracherkennungs-Prozess (209) zugeführt werden, welcher beim Erkennen eines Schlüsselworts oder einer Phrase („*keyword phrase*“) mindestens einen primären Spracherkennungs-Prozess (403) startet oder aus einem Ruhezustand aktiviert,

Zwar werden die Audio-Daten, welche in dem Puffer gespeichert sind, einem sekundären Spracherkennungsprozess 209 im Mobilgerät zugeführt, welcher beim Erkennen eines Schlüsselworts oder einer Phrase („*keyword phrase*“) einen primären Spracherkennungs-Prozess („*speech recognition engine 403*“) startet oder aus einem Ruhezustand aktiviert (Absatz 0038: „*In the idle 303 state, the microcontroller 205*

can receive an indication 215 that a keyword phrase was spoken by the user 101 as detected by the speech recognition 209 circuit. The microcontroller 205 will wake up from its sleep and enter the keyword recognized 312 state.“ und Absatz 0048: „the speech recognition 403 engine which can be a high end, large vocabulary recognizer. Based on the phrase spoken, the computer 402 will execute a predefined function 511, 512, 513“).

Jedoch erfolgt die Zuführung der Audio-Daten an den sekundären Spracherkennungs-Prozess 209 nicht zeitnah im Sinne des Streitpatents, da diese dort direkt nach der Digitalisierung im ADC 5 erfolgt (vgl. Streitpatentschrift, insbesondere Figur 2, Bezugszeichen 11) und nicht erst nach einer Zwischenspeicherung im Audio-Puffer 6, der dort lediglich dem primären Spracherkennungs-Prozess 8 und nicht dem sekundären Spracherkennungsprozess 7 vorgeschaltet ist. Dagegen erfolgt gemäß D4 die Audio-Daten-Übermittlung an den sekundären Spracherkennungsprozess 209 erst nach Zwischenspeicherung im Audio-Puffer („speech capture buffer 204“), sofern dieser überhaupt von der Sprachenergieerkennungsschaltung 203 im betrachteten Zeitraum aktiviert wurde (siehe oben zu Merkmal 1.3), vgl. dazu insbesondere die folgenden Ausschnitte aus den jeweiligen Figuren 2:



Ausschnitt aus Figur 2 der Druckschrift D4 und Ausschnitt aus Figur 2 der Streitpatentschrift mit Ergänzung durch den Senat

- 1.5 d) dass der primäre Spracherkennungs-Prozess (403) den gesamten oder den jüngsten Inhalt des Audio-Puffers („*speech capture buffer 204*“) sowie die sich daran anschließende Liveübertragung in Text umwandelt und diesen Text mindestens einem Dialogsystem-Prozess zuführt,

Der primäre Spracherkennungs-Prozess 403 der D4 wandelt die Audio-Daten in Text um (Absatz 0043: „*The speech recognition module 403 is used to convert received speech input into text as required during the execution of a function.*“).

Voraussetzung für die Umwandlung des Inhalts des Audio-Puffers („*speech capture buffer 204*“) in Text ist, dass der sog. Sprachpfad („*speech path*“) zwischen dem drahtlosen Kommunikationsgerät und der Basis eingerichtet ist (Absatz 0039: „*When the microcontroller 205 initiates a connect request 305 to the base 104, ... speech path is established. In the next step 306, the contents of the speech capture buffer 204 is sent via 214 and 216 to the base 104 for processing, as needed.*“). Die Umwandlung der sich daran anschließenden mit dem Mikrofon aufgenommene Liveübertragung in Text findet statt, wenn der Sprachpfad vollständig eingerichtet ist (Absatz 0039: „*Finally, the microcontroller 205 enters into the connected state 308 thus completing the speech path from the base 104 to the wireless communicator microphone 201 and speaker 208 via paths 217 and 218, respectively.*“).

Dieser Text wird einem anspruchsgemäßen Dialogsystem-Prozess zugeführt (vgl. Merkmale 1.6 und 1.6.1), der ausgeführt wird, wenn die Spracherkennungs-Engine 403 als Erkennungsprogramm mit großem Wortschatz eine vordefinierte Funktion 511 ausführt, bei der es sich beispielsweise um eine Funktion handeln kann, die den Benutzer zu weiteren Eingaben auffordert und somit einen Dialog durchführt (Absatz 0046: „*A base function can be started up 515, .... such that it*

*initiates a base connect request 505 with the wireless communicator 102. Once established, the function can proceed to have a dialog with the user 101 ...” und Absatz 0048: „In the connected 506 state, the computer 402 can engage the speech recognition 403 engine which can be a high end, large vocabulary recognizer. Based on the phrase spoken, the computer 402 will execute a predefined function 511, 512, 513 or resume executing a predefined function already in progress. Each predefined function can request to play an audio prompt or tone to the user to solicit further input from the user 102.” sowie Absatz 0043: „Audio prompts and tone store 406 contains speech prompts or tone indicators that can be played to the user 101, through the speaker of the wireless communicator 102, as necessary in the execution of a function.”).*

- 1.6 welcher ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei,

Der Dialogsystem-Prozess, der bei vollständiger Einrichtung des Sprachpfads gestartet wird (vgl. die oben zum Merkmal 1.5 genannten Textstellen), leitet die mit dem drahtlosen Kommunikationsgerät 102 aufgenommenen Sprachdaten zur Analyse an die Basis 104 weiter (Absatz 0031: „*The purpose of the speech capture buffer 204 is to permit, beyond just keyword speech recognition, ... to have the captured speech be forwarded to the base 104 for analysis or use.*“). Die als Erkennungsprogramm mit großem Wortschatz fungierende Spracherkennungs-Engine 403 analysiert den aus den Sprachdaten gewonnenen Text zur Erkennung gesprochener Phrasen, um eine entsprechende Antwort oder Aktion auszulösen, wobei der Fachmann mitliest, dass dabei ermittelt wird, ob eine Frage, eine Mitteilung und/oder eine Aufforderung ent-

halten ist (Absatz 0048: „*In the connected 506 state, the computer 402 can engage the speech recognition 403 engine which can be a high end, large Vocabulary recognizer. Based on the phrase spoken, the computer 402 will execute a predefined function 511, 512, 513 or resume executing a predefined function already in progress.*” und Anspruch 1, letzter Gliederungspunkt i): „*a speech recognition engine used to recognize any spoken phrases by the user or other speech source*“).

- 1.6.1 falls dies bejaht wird, der Dialogsystem-Prozess eine passende Aktion auslöst oder eine passende Antwort generiert und mit dem Benutzer per Ausgabevorrichtung in Kontakt tritt und

Falls bei der Analyse eine entsprechende Frage, Mitteilung und/oder Aufforderung erkannt wird, löst der Dialogsystem-Prozess eine passende Aktion aus und generiert eine passende Antwort (Absatz 0048: „*Each predefined function can request to play an audio prompt or tone to the user to solicit further input from the user 102. The computer 402 can exchange function-related data with the wireless communicator 102 via a wireless data transfer channel 514 supported by the radio interface 401.*”) und tritt mit dem Benutzer per Ausgabevorrichtung in Kontakt (Absatz 0043: „*Audio prompts and tone store 406 contains speech prompts or tone indicators that can be played to the user 101, through the speaker of the wireless communicator 102*“).

- 1.6.2<sup>Teil</sup> e) dass anderenfalls, falls in den Audio-Daten nur Störgeräusche vorliegen oder falls der Text keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre Spracherkennungs-Prozess (403) und der Dialogsystem-Prozess ~~umge-~~

und wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess (209) zurückgeben.

Im anderen Fall, wenn in den Audio-Daten kein relevanter oder auswertbarer Inhalt enthalten ist, beispielsweise, wenn nur Störgeräusche vorliegen, wird der drahtlose Sprachpfad geschlossen und das drahtlose Kommunikationsgerät und auch der Computer deaktiviert, um den Energieverbrauch zu reduzieren (Absatz 0047: „*The computer 402, when in the connected 506 state, may determine that no speech data is being transferred between the wireless communicator 102 and the base 104. It can signal the wireless communicator 102 to enter into the suspend state and thus greatly reduce battery drain during long silences in the conversation by turning off the wireless speech path. The computer 402 will enter the suspend 507 state*”).

In diesem Ruhezustand kehrt sowohl der primäre Spracherkennungs-Prozess (403) als auch der Dialogsystem-Prozess wieder in den Ruhezustand (303) zurück und die Kontrolle wird wieder allein dem sekundären Spracherkennungs-Prozess (209) zurückgeben, der auf Schlüsselworte bzw. Phrasen wartet (Absatz 0038: „*In the idle 303 state, the microcontroller 205 can receive an indication 215 that a keyword phrase was spoken by the user 101 as detected by the speech recognition 209 circuit. The microcontroller 205 will wake up from its sleep and enter the keyword recognized 312 state.*”). Allerdings erfolgt die Rückkehr in den Ruhezustand nicht umgehend im Sinne des Streitpatents, da diese von einer Kommunikation zwischen dem drahtlose Kommunikationsgerät und der Basis abhängt und auf eine positive Bestätigung des drahtlosen Kommunikationsgeräts



(102) gewartet wird (Absatz 0047: „*The computer 402 will enter the suspend 507 state upon positive acknowledgment from the wireless communicator 102.*“).

Da die Druckschrift **D4** somit die **Merkmale 1.3, 1.4 und 1.6.2** zumindest nicht vollständig offenbart, ist das Verfahren des Patentanspruchs 1 nach Streitpatent gegenüber dem Stand der Technik nach der Druckschrift D4 **neu**.

**2.3** Das Verfahren gemäß Patentanspruch 1 nach Streitpatent beruht gegenüber dem Stand der Technik nach der Druckschrift **D4** auch auf einer **erfinderischen Tätigkeit**.

Wie oben in Abschnitt 2.2 zum **Merkmal 1.3** ausgeführt, werden gemäß dem Verfahren der Druckschrift D4 die mit dem Mikrofon 201 aufgenommenen Audio-Daten nicht kontinuierlich im Audio-Puffer („*speech capture buffer 204*“) zwischengespeichert, so dass dieser nicht stets die Audio-Daten der jüngsten Vergangenheit enthält, da eine Zwischenspeicherung nur dann stattfindet, wenn die Sprachenergieerkennungsschaltung 203 im kontinuierlich empfangenen und abgetasteten binären Datenstrom die Sprachenergie als den vordefinierten Schwellenwert übersteigend bestimmt und den Audio-Puffer aktiviert (Figur 2 i. V. m. Absatz 0026 sowie Anspruch 1, Gliederungspunkte o) bis q)). Eine Veranlassung für den Fachmann den vordefinierten Sprachenergie-Schwellenwert permanent auf den Wert Null einzustellen oder – was dieser Maßnahme entspricht – die Sprachenergieerkennungsschaltung 203 aus dem drahtlosen Kommunikationsgerät zu entfernen, um den fehlenden Teil von Merkmal 1.3 zu realisieren, ist nicht nur nicht erkennbar, sondern bei der Funktionsweise des Systems der D4 nicht praktikabel ohne deren technischen Lehre vollständig zu verwerfen. Denn das mit der technischen Lehre der D4 verfolgte Ziel, die Notwendigkeit eines häufigen Aufladens der Batterie zu minimieren (Absatz 0015), basiert auf dem durch die Sprachenergieerkennungsschaltung 203 veranlassten Ruhezustand des drahtlosen Kommunikationsgeräts 102 (Absatz 0027).

Ebenso hat der Fachmann nicht nur keine Veranlassung den fehlenden Teil von **Merkmal 1.4** zu realisieren, wonach die Audio-Daten zeitnah mindestens einem sekundären Spracherkennungsprozess zugeführt werden, sondern eine derartige zeitnahe Zuführung im Sinne des Streitpatents (vgl. Abschnitt 5.2) würde eine grundlegende Abänderung der technischen Lehre der D4 und der Komponenten des drahtlosen Kommunikationsgeräts 102 und deren Funktionsweise erfordern. Wie oben in Abschnitt 2.2 zum Merkmal 1.4 ausgeführt, findet die Audio-Daten-Übermittlung an den sekundären Spracherkennungsprozess 209 erst nach Zwischenspeicherung im Audio-Puffer („*speech capture buffer 204*“) statt, sofern dieser überhaupt von der Sprachenergieerkennungsschaltung 203 aktiviert wurde (vgl. vorangehenden Absatz), wohingegen die zeitnahe Zuführung der Audio-Daten 11 an den sekundären Spracherkennungsprozess 7 im Sinne des Streitpatents direkt, d. h. ohne die Zwischenschaltung weiterer Komponenten und Schritte erfolgt (vgl. Ausschnitt der jeweiligen Figuren 2 in Abschnitt 2.2). Ein einfaches Verzicht auf den Audio-Puffer (204) vor der Spracherkennungsschaltung 209 in der Anordnung der Druckschrift D4 ist nicht möglich, da jegliche Audio-Daten-Übermittlung ausschließlich über den Audio-Puffer (204) erfolgt und so das gesamte System nicht mehr funktionsfähig wäre. Somit wäre zum Erreichen einer zeitnahen Zuführung der Audio-Daten an den sekundären Spracherkennungsprozess 209 im Sinne des Merkmals 1.4 eine Änderung der Signalpfade und damit eine konzeptionell grundlegende Abänderung der Struktur und Funktionsweise nötig, zu der der Fachmann weder der Druckschrift D4 noch dem übrigen Stand der Technik einen Hinweis oder eine Anregung entnehmen kann.

Schließlich ist es für den Fachmann nicht naheliegend, den fehlenden Teil von **Merkmal 1.6.2** – die Rückkehr bzw. Beenden vom primären Spracherkennungs- und Dialogsystem-Prozess und die Kontrollrückgabe an den sekundären Spracherkennungs-Prozess – ausgehend von Druckschrift D4 als umgehend im Sinne des Streitpatents, d. h. im Wesentlichen ohne Zwischenschritte vorzusehen. Dies liegt darin begründet, dass die Verzögerung keine optionale Maßnahme darstellt, sondern durch die Konzeption des Gesamtsystems bedingt ist.

Die „Rückkehr in den Ruhezustand“ dieser auf dem Computer 402 der Basis ausgeführten Prozesse erfolgt gemäß D4 nicht umgehend im Sinne des Streitpatents,

da diese eine entsprechende Kommunikation zwischen dem drahtlosen Kommunikationsgerät und der Basis erfordert und zusätzlich von einer positiven Bestätigung des drahtlosen Kommunikationsgeräts 102 abhängt (Absatz 0047: „*The computer 402 will enter the suspend 507 state upon positive acknowledgment from the wireless communicator 102.*“), wobei im Falle eines Nicht-Empfangs der Bestätigung seitens des drahtlosen Kommunikationsgeräts 102 die Rückkehr unterbleibt (Absatz 0047: „*If it receives a negative acknowledgment, then the computer 402 will remained in the connected 506 state*“). Dabei handelt es sich somit um ein spezifisches Konzept der Prozessterminierung zur Ressourcenschonung einerseits und zur Vermeidung von überflüssigen Zyklen andererseits, welches grundsätzlich anders gestaltet ist als beim Streitpatent, wo die Rückkehr bei entsprechenden Audio-Daten bzw. Textinhalt umgehend und immer ohne Abhängigkeit von weiteren Bedingungen erfolgt.

Einen Hinweis oder eine Veranlassung für den Fachmann, die in sich geschlossene technische Lehre der Druckschrift D4 zu verwerfen und stattdessen das Konzept des Streitpatents mit den Maßnahmen der fehlenden Teilmerkmale 1.3, 1.4 und 1.6.2 zu realisieren, ist weder in der Druckschrift D4 selbst noch in Kombination mit dem anderen im Verfahren befindlichen Stand der Technik zu erkennen.

Der Fachmann kann somit auch ausgehend von der Druckschrift D4 nicht zum Gegenstand des Patentanspruchs 1 gelangen, ohne erfinderisch tätig zu werden.

**3.** Das Verfahren des erteilten Patentanspruchs 1 ist **neu** gegenüber dem Stand der Technik nach der europäischen Patentschrift **D5 (EP 1 091 346 B1)**, im Folgenden zitiert nach der Übersetzung dieser Patentschrift in Form der Druckschrift **D5a (DE 600 10 827 T2)**, da insbesondere die Merkmale M1.4, 1.5 und M1.6.2 nicht vollständig aus D5/D5a entnehmbar sind. Ausgehend von diesem Stand der Technik ist das Verfahren des Patentanspruchs 1 dem Fachmann weder allein in Verbindung mit seinem Fachwissen noch in der Zusammenschau mit einer oder mehreren der weiteren Druckschriften nahegelegt, so dass es auch auf einer **erfinderischen Tätigkeit beruht**.

**3.1** Die Druckschrift **D5a** betrifft im Allgemeinen das Gebiet von Datenverarbeitungssystemen mit mehreren Quellen („*multi-source data processing systems*“) und beschreibt insbesondere ein sog. Hintergrundsystem zur Wiedergewinnung von Audiodaten für Systeme/Software zur Spracherkennung (Absatz 0001). Das Hintergrundsystem soll mit einem Rechnersystem eingesetzt werden können, welches diktierter Sprache erfasst und verarbeitet, die generiert wird, während die Spracherkennungs-Software einem inaktiven Status zugewiesen ist (Absatz 0018).

**3.2** Die Druckschrift **D5a** offenbart Verfahren zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus, wobei Audio-Daten mit mindestens einem Mikrofon (161) aufgenommen werden (z. B. Absatz 0019 „... *ein Hintergrundsystem zur Wiederherstellung von Audiodaten, das ein Programm-Modul zur Spracherkennung ausweist, Audiodaten aufzeichnen kann und anschließend Spracherkennungs-Techniken auf die im Hintergrund erfassten Sprachdaten oder die Audiodaten anwendet, die von einem Mikrofon empfangen wurden, wenn dem Programm-Modul zur Spracherkennung versehentlich ein inaktiver Modus zugewiesen wurde. ... verhindert den Verlust eines Diktats von einem Benutzer.*“) gemäß den **Merkmale 1.1 und 1.2**.

Die Audio-Daten werden in mindestens einem Audio-Puffer zwischengespeichert (Absatz 0020: „*Wie oben angegeben, speichert das Hintergrundsystem zur Wiedergewinnung von Audiodaten der vorliegenden Erfindung kontinuierlich Sprache oder Audiodaten im Hintergrund*“ und Absatz 00027: „*Wenn Audiodaten von der Audio-daten-Eingangsvorrichtung (d.h. einem Mikrofon) empfangen werden, speichert das Hintergrundsystem zur Wiedergewinnung von Audiodaten die Audiodaten in einer Speichervorrichtung.*“ i. V. m. Figur 2 sowie Absatz 0089 i. V. m. den Figuren 10 und 11). Die Zwischenspeicherung erfolgt auch – entgegen der Ansicht der Beklagten – kontinuierlich, so dass der Audio-Puffer stets die Audio-Daten der jüngsten Vergangenheit enthält, (vgl. die senatsseitigen Unterstreichungen im vorherigen Zitat), weshalb auch **Merkmal 1.3** aus D5a bekannt ist.

Hinsichtlich der **Merkmale 1.4 und 1.5** ist der D5a nicht eindeutig entnehmbar, ob zusätzlich zu dem aktivierten primären Spracherkennungs-Prozess des Programm-

Moduls zur Spracherkennung 240 ein zweiter separater sekundärer Spracherkennungs-Prozess existiert, der Schlüsselworte oder Phrasen erkennt. Denn das Hintergrundsystem zur Wiedergewinnung von Audiodaten 10, das von der Klägerin als sekundärer Spracherkennungsprozess identifiziert wird, stellt das übergeordnete Gesamtsystem dar (vgl. z. B. die Bezugszeichen 10 und 240 in Figur 3), und Absatz 0090 besagt: „Das Hintergrundsystem zur Wiedergewinnung von Audiodaten 10 überwacht ... jeden gesprochenen Befehl, der durch das Programm-Modul zur Spracherkennung 240 verarbeitet wird.“. Daher sind die **Merkmale 1.4 und 1.5** zumindest **nicht vollständig und eindeutig** aus der D5a zu entnehmen.

Die **Merkmale 1.6 und 1.6.1**, wonach der Dialogsystem-Prozess ebenfalls startet oder aus dem Ruhezustand aktiviert wird und den Inhalt des Textes darauf hin analysiert, ob dieser eine Frage, eine Mitteilung und/oder eine Aufforderung enthält, die vom Benutzer an den Software-Agenten gerichtet wurde, wobei, falls dies bejaht wird, der Dialogsystem-Prozess eine passende Aktion auslöst oder eine passende Antwort generiert und mit dem Benutzer per Ausgabevorrichtung in Kontakt tritt, sind in der Druckschrift D5a offenbart (Absätze 0095 und 0096 sowie Figur 13, Bezugszeichen 404 und 408).

Jedoch ist **Merkmal 1.6.2**, wonach, falls in den Audio-Daten nur Störgeräusche vorliegen oder falls der Text keinen relevanten oder keinen auswertbaren Inhalt enthält, der primäre Spracherkennungs-Prozess und der Dialogsystem-Prozess umgehend wieder in den Ruhezustand zurückkehren oder sich beenden und die Kontrolle wieder dem sekundären Spracherkennungs-Prozess zurückgeben, der Druckschrift D5a **nicht entnehmbar**.

Da die Druckschrift **D5a** somit **Merkmal 1.6.2** nicht und die **Merkmale 1.4 und 1.5** zumindest nicht vollständig und eindeutig offenbart, ist das Verfahren des Patentanspruchs 1 nach Streitpatent gegenüber dem Stand der Technik nach der Druckschrift D5/D5a **neu**.

**3.3** Das Verfahren gemäß Patentanspruch 1 nach Streitpatent **beruht** gegenüber dem Stand der Technik nach der Druckschrift **D5/D5a** auch auf einer **erfindrischen Tätigkeit**.

Die Druckschrift D5a stellt bereits keinen geeigneten Ausgangspunkt für den Fachmann dar. Der Fachmann, der ein Verfahren schaffen will, mit dem es möglich ist, einem Software-Agenten, der sich in einem energiesparenden Standby- bzw. Schlafzustand befindet, per „natürlicher“ Sprache komplexe Fragen zu stellen, wobei das System umgehend mit einer vollständigen Antwort oder mit einer Aktion reagieren soll (vgl. Streitpatentschrift, Absatz 0009), wird kein Dokument heranziehen, welches darauf ausgerichtet ist, die Diktierfunktion eines Textverarbeitungsprogramms wie WORD zu verbessern, indem bei versehentlich nicht aktivierter Sprachaufnahme verlorener Text wiederhergestellt werden kann.

Doch selbst wenn der Fachmann von der Druckschrift D5a ausgehen würde, hätte er keine Veranlassung, deren abgeschlossene technische Lehre so abzuändern, dass falls die Audio-Daten keinen relevanten oder auswertbaren Inhalt enthalten, den Spracherkennungsprozess und den Dialogsystem-Prozess umgehend wieder in den Ruhezustand zurückkehren zu lassen oder zu beenden. Diese Maßnahme dient offensichtlich – und wie das Streitpatent mehrfach betont – dazu, Energie zu sparen. Die Frage der „Energieeffizienz“ wird in D5a nachvollziehbarerweise nicht thematisiert, da das dortige Verfahren (Absatz 0056: „*ein Textverarbeitungsprogramm 210 (wie zum Beispiel WORD ...)*“) in der Regel auf einem Desktop-PC oder Laptop (z. B. Absatz 0048: „*Personal-Computer*“) und nicht auf einem Smartphone oder dergleichen ausgeführt wird, weshalb die Energieeffizienz keine oder eine nur sehr untergeordnete Rolle spielt.

Da der Fachmann somit weder in der D5a selbst noch in Kombination mit dem anderen vorliegenden Stand der Technik einen Hinweis oder eine Anregung bekommt, das fehlende Merkmal 1.6.2 zu realisieren, kann das Verfahren nach Anspruch 1 ausgehend von der Druckschrift D5a nicht nahegelegt werden.

**4.** Auch die Druckschriften **D1 (WO 2012/025784 A1)** und **D3 (US 7,424,431 B2)** vermögen die Neuheit und erfinderische Tätigkeit des Gegenstands von Patentanspruch 1 nach Hauptantrag nicht in Frage zu stellen, liegen

weiter ab und bilden bei der Beurteilung der erfinderischen Tätigkeit keinen geeigneten Ausgangspunkt für ein Naheliegen, wie bereits im qualifizierten Hinweis des Senats ausgeführt wurde und insoweit von der Klägerin danach auch nicht mehr aufgegriffen worden ist.

**4.1** Zwar mag es ausgehend von der Lehre der Druckschrift **D1** für den Fachmann noch naheliegend sein, den **fehlenden Teil des Merkmals 1.6.2.** zu realisieren. Denn bei dem in D1 vorgeschlagenen Verfahren wird bereits davon ausgegangen, dass der Prozessor bei einer herkömmlichen Spracherkennungsfunktion zu viel Energie verbraucht, um die ganze Zeit eingeschaltet zu sein, da die Energiezelle der jeweiligen elektronischen Vorrichtung, in der die Audio-Benutzerschnittstelle eingebaut ist, zu schnell entladen werden würde (Seite 1, letzter Absatz: „...*drain a power cell of the electronic device too quickly.*“). Obwohl dieses Problem durch die in der D1 vorgeschlagenen Lösung, wonach die Audio-Benutzerschnittstelle erst aktiviert wird, wenn ein Schlüsselwort erkannt wird, bereits signifikant vermindert wird, könnte der Fachmann dennoch veranlasst sein, es nicht allein bei dieser ressourcensparenden Maßnahme zu belassen, sondern den energieintensiven Prozessor und damit auch den primären Spracherkennungs- und den Dialogsystem-Prozess nicht erst nach einer Zeitüberschreitung (Seite 18, Zeilen 1 bis 3: „*When the audio user interface has timed out, then the processor 110 can return to a sleep mode ...*“), sondern „umgehend“ zu deaktivieren bzw. in den Ruhezustand zurückzusetzen im Sinne des Merkmals 1.6.2.

Jedoch handelt es sich bei dem Unterschied zwischen dem aus der Druckschrift D1 bekannten Verfahren und dem Verfahren gemäß dem Patentanspruch 1 nach Streitpatent mit dem in D1 **fehlenden Teil des Merkmals 1.5**, wonach nicht nur die Liveübertragung, sondern auch der gesamte oder der jüngste Inhalt des Audio-Puffers vom primären Spracherkennungs-Prozess einem Dialogsystem-Prozess zugeführt wird, um die Verwirklichung eines grundsätzlich anderen Konzepts.

Es sind keine Anhaltspunkte ersichtlich – weder in der Druckschrift D1 selbst noch in Kombination mit dem anderen im Verfahren befindlichen Stand der Technik – die den Fachmann veranlassen könnten, einen Audio-Puffer vorzusehen, um die darin

schon vor der Erkennung eines Schlüsselwortes gespeicherten Audiodaten zu analysieren, für einen Software-Agenten zu verwenden und damit das fehlende Teilmerkmal 1.5 zu realisieren. Vielmehr würde eine solche Überlegung ausgehend von der Druckschrift D1 nach Überzeugung des Senats eine unzulässige rückschauende Betrachtungsweise in Kenntnis der technischen Lehre des Streitpatents darstellen.

**4.2** Die technische Lehre der Druckschrift **D3** kann schon keinen geeigneten Ausgangspunkt für einen Fachmann darstellen, der das Ziel hat, ein Verfahren bereitzustellen, mit dem es möglich ist, einem Software-Agenten, der sich in einem Standby-Zustand befindet, per „natürlicher“ Sprache komplexe Fragen zu stellen, auf die das System umgehend mit einer vollständigen Antwort oder mit einer Aktion reagiert, so dass für den Benutzer der Unterschied zwischen Standby-Zustand und dem regulären Betrieb kaum bemerkbar ist. Denn die dazu zwingend erforderliche kontinuierliche Zwischenspeicherung von Audio-Daten der jüngsten Vergangenheit in einem Audio-Puffer ist in der Druckschrift D3 nicht bekannt (**Merkmal 1.3 fehlt**), weshalb auch keine Umwandlung von in einem Audio-Puffer gespeicherten Inhalt erfolgen kann (**Teil von Merkmal 1.5 fehlt**).

Falls der Fachmann trotzdem vom Verfahren der Druckschrift D3 ausgehen sollte, könnte es für ihn zwar noch naheliegend sein, den primären Spracherkennungsprozess und den Dialogsystem-Prozess nicht erst nach einer Zeitüberschreitung, sondern „umgehend“ im Sinne des fehlenden **Merkmals 1.6.2** zu beenden, da auch ein energiesparendes Betreiben des Media Players in der Druckschrift D3 thematisiert wird (Spalte 4, Zeilen 22 bis 30: „to power down ... low power mode“).

Jedoch ist kein Hinweis oder eine Veranlassung – weder in der Druckschrift D3 selbst noch in Kombination mit dem anderen im Verfahren befindlichen Stand der Technik – für den Fachmann zu erkennen, das dort beschriebene Verfahren in Richtung auf die **Merkmale 1.3 und 1.5** des Patentanspruchs 1 nach Streitpatent abzuändern. Denn zum einen würde das zusätzliche Einfügen der Hardware-Komponente eines Audio-Puffers zusammen mit zeitgesteuertem Abspeichern und Auslesen eine grundsätzliche und aufwändige Umgestaltung des Verfahrens bedeuten.



Zum anderen dürfte für die Steuerung eines Medienabspielgeräts mit einer festgelegten und wenig komplexen Menüstruktur sowie einer nicht zeitkritischen Anforderung an die Reaktion bei Befehlseingaben ein derartiger Mehraufwand unnötig sein.

5. Noch weiter entfernt von dem nach Anspruch 1 des Streitpatents beanspruchten Verfahren liegt die technische Lehre der Druckschrift **D6 (US 7,610,199 B2)**, weshalb auch dieser Stand der Technik dessen Neuheit und Beruhen auf erfinderischer Tätigkeit nicht in Frage stellen kann.

5.1 Die Druckschrift **D6** beschreibt ein Verfahren und eine Vorrichtung zur Gewinnung vollständiger Sprachsignale für Spracherkennungsanwendungen. In einer Ausführungsform zeichnet das Verfahren kontinuierlich einen Audiostrom in einem Ringpuffer auf. Wenn ein Benutzerbefehl zum Starten oder Beenden der Spracherkennung empfangen wird (Spalte 3, Zeilen 7 bis 10 und Zeilen 33 bis 35: *„user command (e. g., via a button press or other means)...“*), erhält das Verfahren eine Anzahl von Rahmen (*„frame“*) des Audiostroms, die vor oder nach dem nach dem Benutzerbefehl auftreten, um ein erweitertes Audiosignal für die Spracherkennungsverarbeitung zu identifizieren. In weiteren Ausführungsformen analysiert das Verfahren das erweiterte Audiosignal, um Anfangs- und Endpunkte der Sprache zu lokalisieren, die zumindest einen Teil der zu verarbeitenden Sprache für die Erkennung festlegen. Mindestens einer der Sprachendpunkte wird mit Hilfe eines Hidden Markov Modells lokalisiert (Abstract). Die Druckschrift D6 befasst sich also mit der mangelhaften Synchronisation zwischen Tastendruck und Sprachäußerungen seitens des Anwenders (Spalte 2, Zeilen 53 bis 61: *„user commands that are not properly synchronized with user utterances.“*).

5.2 Aus Druckschrift **D6** ist entnehmbar, dass ein Software-Agent aus einem Standby-Modus aktiviert wird (Spalte 2, Zeilen 9 bis 20: *„the present invention relates to a method and apparatus for obtaining complete speech signals for speech recognition applications. ... When a user command to commence or terminate speech recognition is received; **Merkmal 1.1)**, Audio-Daten mit einem Mikrofon aufgenommen werden (Spalte 3, Zeilen 25 bis 27: *„the audio signal for speech**

*recognition processing has been acquired using an open-microphone mode“; **Merkmal 1.2)** und kontinuierlich in einem Audio-Puffer zwischengespeichert werden, so dass dieser die Audio-Daten der jüngsten Vergangenheit enthält (Spalte 3, Zeilen 3 bis 7: „where the method 100 continuously records an audio stream (e. g., a sequence of audio frames containing user speech, background audio, etc.) to a circu- lar buffer.“ i. V. m. Figur 1, Bezugszeichen 104; **Merkmal 1.3)**.*

In Druckschrift D6 ist jedoch **Merkmal 1.4 nicht** vollständig offenbart, da, unabhängig von der Frage, ob sich zwei Spracherkennungs-Prozesse identifizieren lassen, das Aktivieren des Spracherkennungsprozesses in D6 nicht durch Erkennen eines Schlüsselworts oder einer Phrase in den Audio-Daten erfolgt, sondern es wird dort explizit lediglich als manuell durch Tastendruck ausgelöst beschrieben. Ansonsten wird der „*command*“ für den Spracherkennungsprozess nur pauschal als durch andere Mittel realisierbar angegeben (Spalte 1, Zeilen 38 bis 43: „*For instance, a user may begin speaking before he provides the command to process his speech (e.g., by pressing a button), or he may terminate the processing command before he is finished uttering the speech signal to be processed (e.g., by releasing or pressing a button).*“ und Spalte 3, Zeilen 7 bis 10: „*a user command (e.g., via a button press or other means) to commence speech recognition*“ sowie Spalte 3, Zeilen 33 bis 35: „*a user command (e.g., via a button press or other means) to terminate speech recognition*“).

Da in D6 zudem auch die Ausgestaltung und Funktionsweise der einem primären Spracherkennungsprozesses entsprechenden Spracherkennung gemäß **Merkmal 1.5 nicht** entnehmbar ist, ist zu den darauf aufbauenden Verfahrensschritten der **Merkmalsgruppe 1.6** lediglich noch der **Teil von Merkmal 1.6.2** eindeutig offenbart, dass die Beendigung des Spracherkennungsprozesses erfolgt, wenn keine Sprache mehr vom System erkannt wird.

**5.3** Obwohl in Druckschrift **D6** angegeben wird, dass die vorgeschlagene Spracherkennungsverarbeitung in Übereinstimmung mit jeder bekannten Spracherkennungstechnik angewandt werden kann (Spalte 3, Zeilen 57 bis 60), stellt die technische Lehre mit einem manuellen Aktivieren des Spracherkennungsprozesses

durch Tastendruck gerade keinen geeigneten Ausgangspunkt für den Fachmann dar, der ein Verfahren zur Sprachaktivierung per „natürlicher“ Sprache ohne haptische Bedienung bzw. ohne einen Touchscreen zu berühren oder eine Taste zu drücken schaffen will (vgl. Streitpatentschrift, Absätze 0002 und 0006).

Doch selbst wenn der Fachmann von der Druckschrift **D6** ausgehen würde, hätte er keine Veranlassung die in sich geschlossene Lehre der D6 so abzuändern, dass ein zweistufiges Spracherkennungsprozess-Verfahren mit Dialogsystem, Starten durch Schlüsselwort- oder Phrasenerkennung und verbesserter Energieeffizienz realisiert würde, da er zu einer so umfangreichen Änderung weder aus der D6 Anregungen oder Hinweise bekommt, noch in Kombination mit den Druckschriften D1 und/oder D3 – wie die Klägerin meint – oder einer der anderen Druckschriften aus dem vorliegenden Stand der Technik.

**6.** Die übrigen Druckschriften **D7 (US 7,610,199 B2)**, **D8 (US 2009/0204410 A1)** und **D9 (EP 1 091 347 A2)** liegen noch weiter vom Streitpatentgegenstand ab und wurden von der Klägerin lediglich in das Verfahren eingeführt, um zu belegen, dass dem Fachmann die Korrektur von Falsch-Positiv-Fehlern eines Spracherkennungsprozesses durch einen anderen Spracherkennungsprozess bekannt gewesen sei. Dass weitergehende Merkmale des Streitpatents diesen Druckschriften zu entnehmen sind, wurde nicht geltend gemacht und ist auch nicht ersichtlich.

**7.** Da sich mithin der Gegenstand des Patentanspruchs 1 in seiner erteilten Fassung nach Hauptantrag für den Fachmann nicht in naheliegender Weise aus dem im Verfahren befindlichen Stand der Technik ergibt, gilt er als auf einer erfinderischen Tätigkeit beruhend und ist patentfähig.

**8.** Die vorstehenden Ausführungen zur Patentfähigkeit des Verfahrens gemäß dem Anspruchs 1 nach Hauptantrag gelten entsprechend auch für den erteilten Anspruch 12, dessen Gegenstand ein System zur Sprachaktivierung eines Software-Agenten aus einem Standby-Modus betrifft und inhaltlich die gleichen im Anspruch 1 genannten Merkmale enthält.

Die ebenfalls angegriffenen Unteransprüche 2 bis 11 und 13 bis 15, welche vorteilhafte Ausgestaltungen des Erfindungsgegenstands betreffen, sind bereits durch ihren Rückbezug auf die jeweils patentfähigen Ansprüche 1 bzw. 12 ebenfalls rechtsbeständig. Gegenteiliges hat auch die Klägerin weder geltend gemacht noch dargelegt.

Auf die Hilfsanträge kam es daher nicht an.

## **B. Nebenentscheidungen**

Die Kostenentscheidung beruht auf § 84 Abs. 2 PatG i. V. m. § 91 Abs. 1 ZPO.

Die Entscheidung über die vorläufige Vollstreckbarkeit beruht auf § 99 Abs. 1 PatG i. V. m. § 709 ZPO.

## **C. Rechtsmittelbelehrung**

Gegen dieses Urteil ist das Rechtsmittel der Berufung gegeben.

Die Berufungsschrift muss von einer in der Bundesrepublik Deutschland zugelassenen **Rechtsanwältin oder Patentanwältin** oder von einem in der Bundesrepublik Deutschland zugelassenen **Rechtsanwalt oder Patentanwalt** unterzeichnet oder im Fall der elektronischen Einreichung mit einer qualifizierten elektronischen Signatur nach dem Signaturgesetz oder mit einer fortgeschrittenen elektronischen Signatur versehen sein, die von einer internationalen Organisation auf dem Gebiet des gewerblichen Rechtsschutzes herausgegeben wird und sich zur Bearbeitung durch das jeweilige Gericht eignet.

Die Berufungsschrift muss die Bezeichnung des Urteils, gegen das die Berufung gerichtet wird, sowie die Erklärung enthalten, dass gegen dieses Urteil Berufung

eingelegt werde. Mit der Berufungsschrift soll eine Ausfertigung oder beglaubigte Abschrift des angefochtenen Urteils vorgelegt werden.

Die Berufungsschrift muss **innerhalb eines Monats** schriftlich bzw. in elektronischer Form beim Bundesgerichtshof, Herrenstraße 45a, 76133 Karlsruhe eingereicht werden. Die Berufungsfrist beginnt mit der Zustellung des in vollständiger Form abgefassten Urteils, spätestens aber mit dem Ablauf von fünf Monaten nach der Verkündung. Die Frist ist nur gewahrt, wenn die Berufung vor Fristablauf beim Bundesgerichtshof eingeht.

Voit

Altvater

v. Hartz

Haupt

Tischler