

---

## Voice Assistant - Computer Wake-Word Project

Dokument: Wake-Word Testing Framework

Datum: 05. Dezember 2025

---

Seite: {page}

---

# Wake-Word Testing Framework

---

## Systematisches Testen des “Computer” Wake-Words

---

---

### Inhaltsverzeichnis

---

1. [Test-Übersicht](#)
  2. [Erkennungsrate-Tests](#)
  3. [Falsch-Positiv-Tests](#)
  4. [Umgebungs-Tests](#)
  5. [Stress-Tests](#)
  6. [Performance-Tests](#)
  7. [Erfolgs-Kriterien](#)
  8. [Test-Protokoll](#)
- 

### Test-Übersicht

---

Dieses Dokument beschreibt ein systematisches Testing-Framework für das “Computer” Wake-Word. Alle Tests sollten durchgeführt werden, bevor das Wake-Word als produktionsreif betrachtet wird.

## Warum systematisches Testing?

Ein Wake-Word-System muss in verschiedenen Szenarien zuverlässig funktionieren. Systematisches Testing stellt sicher, dass das System nicht nur unter idealen Bedingungen funktioniert, sondern auch in realen Nutzungsszenarien mit Hintergrundgeräuschen, verschiedenen Sprechern und unterschiedlichen Umgebungen.

## Test-Kategorien

Die Tests sind in fünf Hauptkategorien unterteilt:

**Erkennungsrate-Tests** messen, wie zuverlässig das Wake-Word erkannt wird, wenn es tatsächlich gesprochen wurde. Diese Tests sind fundamental, da eine niedrige Erkennungsrate die Nutzbarkeit des Systems stark beeinträchtigt.

**Falsch-Positiv-Tests** messen, wie oft das System fälschlicherweise aktiviert wird, wenn das Wake-Word nicht gesprochen wurde. Zu viele Falsch-Positive führen zu Frustration und reduzieren das Vertrauen in das System.

**Umgebungs-Tests** stellen sicher, dass das System in verschiedenen akustischen Umgebungen funktioniert, von leisen Büros bis zu lauten öffentlichen Räumen.

**Stress-Tests** prüfen die Langzeit-Stabilität und Robustheit des Systems unter kontinuierlicher Nutzung.

**Performance-Tests** messen technische Metriken wie Latenz, CPU-Auslastung und Memory-Verbrauch.

---

## Erkennungsrate-Tests

---

### Test 1.1: Basis-Erkennungsrate

**Ziel:** Messen der grundlegenden Erkennungsrate unter idealen Bedingungen

**Setup:**

- Leise Umgebung (Büro oder Zimmer)

- Normale Sprechlautstärke
- Abstand zum Mikrofon: 1-2 Meter
- Keine Hintergrundgeräusche

### Durchführung:

1. Starte Voice Assistant
2. Sage 100x "Computer" in normalem Tempo
3. Zähle erfolgreiche Erkennungen
4. Dokumentiere Fehlerkennungen

### Checkliste:

Test 1.1: Basis-Erkennungsrate  
 Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Versuch	Erkannt	Notizen
1-10	__ /10	
11-20	__ /10	
21-30	__ /10	
31-40	__ /10	
41-50	__ /10	
51-60	__ /10	
61-70	__ /10	
71-80	__ /10	
81-90	__ /10	
91-100	__ /10	

GESAMT: \_\_\_ /100 (\_\_\_%)

Erfolg: [ ] Ja ( $\geq 95\%$ ) [ ] Nein ( $< 95\%$ )

**Erwartetes Ergebnis:**  $\geq 95\%$  Erkennungsrate

### Test 1.2: Verschiedene Tonlagen

**Ziel:** Testen der Erkennungsrate bei verschiedenen Tonhöhen

**Setup:** Wie Test 1.1

**Durchführung:**

1. 20x normale Stimme
2. 20x höhere Tonlage
3. 20x tiefere Tonlage
4. 20x geflüstert
5. 20x laut

**Checkliste:**

Test 1.2: Verschiedene Tonlagen

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Tonlage	Versuche	Erkannt	Rate
Normal	20	—	__%
Hoch	/ 20	/ —	__%
Tief	/ 20	/ —	__%
Geflüstert	/ 20	/ —	__%
Laut	/ 20	/ —	__%

*GESAMT:* \_\_\_ /100 (\_\_\_%)

*Erfolg:* [ ] Ja (>90%) [ ] Nein (<90%)

**Erwartetes Ergebnis:** ≥90% Erkennungsrate (etwas niedriger als Basis)

## Test 1.3: Verschiedene Geschwindigkeiten

**Ziel:** Testen der Erkennungsrate bei verschiedenen Sprechgeschwindigkeiten

**Durchführung:**

1. 30x normale Geschwindigkeit
2. 30x schnell (“Compu-ter”)
3. 30x langsam (“Com...pu...ter”)

## Checkliste:

Test 1.3: Verschiedene Geschwindigkeiten

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Geschwindigkeit	Versuche	Erkannt	Rate
Normal	30	__	__%
<b>Schnell</b>	/ 30	/ __	/ __%
<i>Langsam</i>	/ 30	/ __	/ __%

**GESAMT:** \_\_\_ /90 (\_\_\_%)

*Erfolg:* [ ] Ja ( $\geq 90\%$ ) [ ] Nein ( $< 90\%$ )

**Erwartetes Ergebnis:**  $\geq 90\%$  Erkennungsrate

---

## Test 1.4: Verschiedene Abstände

**Ziel:** Testen der Erkennungsrate bei verschiedenen Entfernungen zum Mikrofon

### Durchführung:

1. 20x aus 0.5m Entfernung
2. 20x aus 1m Entfernung
3. 20x aus 2m Entfernung
4. 20x aus 3m Entfernung
5. 20x aus 4m Entfernung

### Checkliste:

#### Test 1.4: Verschiedene Abstände

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Abstand	Versuche	Erkannt	Rate
0.5m	20	__	__%
1.0m	/ 20	/ __	/ __%
2.0m	/ 20	/ __	/ __%
3.0m	/ 20	/ __	/ __%
4.0m	/ 20	/ __	/ __%

**GESAMT:** \_\_\_ /100 (\_\_\_%)

**Erfolg:** [ ] Ja ( $\geq 80\%$  bei 1-3m) [ ] Nein

**Erwartetes Ergebnis:**  $\geq 95\%$  bei 1-2m,  $\geq 80\%$  bei 3m

---

## Falsch-Positiv-Tests

---

### Test 2.1: Ähnliche Wörter

**Ziel:** Testen, ob ähnlich klingende Wörter fälschlicherweise erkannt werden

**Durchführung:** Sage jedes Wort 10x und zähle Fehlerkennungen

**Checkliste:**

### Test 2.1: Ähnliche Wörter

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Wort	Versuche	Fehl-Erkannt	Rate
Commuter	10	__	___%
<b>Komputer</b>	/ 10	/ __	/ ___%
<i>Puter</i>	/ 10	/ __	/ ___%
<i>Commute</i>	/ 10	/ __	/ ___%
<i>Compute</i>	/ 10	/ __	/ ___%
<b>Kompute</b>	/ 10	/ __	/ ___%
<i>Rechner</i>	/ 10	/ __	/ ___%
<i>PC</i>	/ 10	/ __	/ ___%
<i>Laptop</i>	/ 10	/ __	/ ___%
<i>Tablet</i>	/ 10	/ __	/ ___%

*GESAMT:* \_\_\_\_ **Fehl-Erkennungen von 100 Versuchen** (\_\_\_\_%)

Erfolg: [ ] Ja (<5% Fehl-Erkennungen) [ ] Nein ( $\geq 5\%$ )

## Erwartetes Ergebnis: % Falsch-Positive

---

## Test 2.2: Normale Konversation

**Ziel:** Testen, ob normale Gespräche Fehlerkennungen auslösen

### Setup:

- Führe normale Gespräche in der Nähe des Mikrofons
- Vermeide absichtlich das Wort “Computer”
- Dauer: 1 Stunde

### Durchführung:

1. Starte Voice Assistant
2. Führe normale Gespräche (oder spiele Podcast/Video ab)
3. Zähle Fehlerkennungen

### Checkliste:

Test 2.2: Normale Konversation  
Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Zeit (Minuten)	Fehl-Erkennungen	Notizen
0-10	__	
10-20	__	
20-30	__	
30-40	__	
40-50	__	
50-60	__	

GESAMT: \_\_\_ Fehl-Erkennungen in 60 Minuten

Erfolg: [ ] Ja (<1 pro Stunde) [ ] Nein ( $\geq 1$  pro Stunde)

**Erwartetes Ergebnis:** Fehl-Erkennung pro Stunde

---

## Test 2.3: Hintergrundgeräusche

**Ziel:** Testen, ob Hintergrundgeräusche Fehlerkennungen auslösen

**Durchführung:** Teste mit verschiedenen Hintergrundgeräuschen (je 10 Minuten)

**Checkliste:**

### Test 2.3: Hintergrundgeräusche

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Geräusch-Typ	Dauer	Fehl-Erkennungen	Rate/h
Musik (leise)	10min	__	__
Musik (laut)	10min	__	__
TV/Radio	10min	__	__
Straßenlärm	10min	__	__
Tippen/Klicken	10min	__	__
Ventilator	10min	__	__

GESAMT: \_\_ Fehl-Erkennungen in 60 Minuten

Erfolg: [ ] Ja (<2 pro Stunde) [ ] Nein ( $\geq 2$  pro Stunde)

**Erwartetes Ergebnis:** Fehl-Erkennungen pro Stunde

---

## Umgebungs-Tests

---

### Test 3.1: Verschiedene Räume

**Ziel:** Testen in verschiedenen akustischen Umgebungen

**Checkliste:**

### Test 3.1: Verschiedene Räume

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Raum-Typ	Versuche	Erkannt	Rate	Notizen
Kleines Zimmer	20	__	__%	/
Großes Zimmer	20	__	__%	/
Büro	20	__	__%	/
Küche	20	__	__%	/
Badezimmer (Hall)	20	__	__%	/
Draußen	20	__	__%	/

GESAMT: \_\_\_ /120 (\_\_\_%)

Erfolg: [ ] Ja ( $\geq 85\%$  in allen Räumen) [ ] Nein

**Erwartetes Ergebnis:**  $\geq 85\%$  in allen Räumen

### Test 3.2: Verschiedene Lautstärke-Umgebungen

**Ziel:** Testen bei verschiedenen Umgebungslautstärken

**Checkliste:**

### Test 3.2: Verschiedene Lautstärke-Umgebungen

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Umgebung	Versuche	Erkannt	Rate	Fehl-Pos
Sehr leise (<30dB)	20	__	__%	__
Leise (30-40dB)	20	__	__%	__
Normal (40-50dB)	20	__	__%	__
Laut (50-60dB)	20	__	__%	__
Sehr laut (>60dB)	20	__	__%	__

GESAMT: \_\_\_ /100 (\_\_\_%)

Erfolg: [ ] Ja ( $\geq 80\%$  bei 30-60dB) [ ] Nein

**Erwartetes Ergebnis:**  $\geq 80\%$  bei normalen Lautstärken (30-60dB)

# Stress-Tests

## Test 4.1: Schnelle Wiederholungen

**Ziel:** Testen bei schnellen aufeinanderfolgenden Aktivierungen

**Durchführung:** Sage “Computer” 100x hintereinander mit nur 2 Sekunden Pause

**Checkliste:**

Test 4.1: Schnelle Wiederholungen			
Datum:	Tester:		
Versuch	Erkannt	Doppel-Erkennung	Verzögerung
1-10	__ /10	__	[ ] Ja/Nein
11-20	__ /10	__	[ ] Ja/Nein
21-30	__ /10	__	[ ] Ja/Nein
31-40	__ /10	__	[ ] Ja/Nein
41-50	__ /10	__	[ ] Ja/Nein
51-60	__ /10	__	[ ] Ja/Nein
61-70	__ /10	__	[ ] Ja/Nein
71-80	__ /10	__	[ ] Ja/Nein
81-90	__ /10	__	[ ] Ja/Nein
91-100	__ /10	__	[ ] Ja/Nein

GESAMT: \_\_ /100 (\_\_\_\_%)  
Doppel-Erkenntnisse: \_\_

Erfolg: [ ] Ja ( $\geq 90\%$ , keine Doppel-Erkennungen) [ ] Nein

**Erwartetes Ergebnis:**  $\geq 90\%$ , keine Doppel-Erkennungen

## Test 4.2: Langzeit-Stabilität

**Ziel:** Testen der Stabilität über längere Zeit

**Durchführung:** Lasse Voice Assistant 1 Stunde laufen und aktiviere alle 5 Minuten

## Checkliste:

### Test 4.2: Langzeit-Stabilität

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Zeit (Min)	Aktivierung	Erkannt	CPU%	Memory(MB)
5	Computer	[ ]	—	—
10	Computer	[ ]	—	—
15	Computer	[ ]	—	—
20	Computer	[ ]	—	—
25	Computer	[ ]	—	—
30	Computer	[ ]	—	—
35	Computer	[ ]	—	—
40	Computer	[ ]	—	—
45	Computer	[ ]	—	—
50	Computer	[ ]	—	—
55	Computer	[ ]	—	—
60	Computer	[ ]	—	—

GESAMT: \_\_\_ /12 (\_\_\_%)

Durchschn. CPU: \_\_\_%

Durchschn. Memory: \_\_\_MB

Erfolg: [ ] Ja ( $\geq 95\%$ , stabil) [ ] Nein

**Erwartetes Ergebnis:**  $\geq 95\%$ , keine Memory Leaks, CPU stabil

## Performance-Tests

### Test 5.1: Latenz-Messung

**Ziel:** Messen der Zeit von Wort-Ende bis Erkennung

## Checkliste:

**Test 5.1: Latenz-Messung**

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Versuch | Latenz (ms) | Notizen

Versuch	Latenz (ms)	Notizen
1	__	
2	__	
3	__	
4	__	
5	__	
6	__	
7	__	
8	__	
9	__	
10	__	

Durchschnitt: \_\_\_ ms

Min: \_\_\_ ms

Max: \_\_\_ ms

Erfolg: [ ] Ja (&lt;100ms Durchschnitt) [ ] Nein

**Erwartetes Ergebnis:** <100ms Durchschnitt, <200ms Maximum

---

**Test 5.2: Ressourcen-Verbrauch****Ziel:** Messen von CPU und Memory Auslastung**Checkliste:**

### Test 5.2: Ressourcen-Verbrauch

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Zustand	CPU%	Memory(MB)	Notizen
Idle (wartend)	—	—	
Wake-Word erkannt	—	—	
STT aktiv	—	—	
TTS aktiv	—	—	
Nach 10min	—	—	
Nach 30min	—	—	
Nach 60min	—	—	

Durchschn. CPU: \_\_\_\_%

*Durchschn. Memory:* \_\_\_\_MB

Erfolg: [ ] Ja (<20% CPU, <200MB Memory) [ ] Nein

**Erwartetes Ergebnis:** <20% CPU, <200MB Memory

---

## Erfolgs-Kriterien

---

### Minimale Anforderungen

Ein Wake-Word-System gilt als produktionsreif, wenn folgende Kriterien erfüllt sind:

#### ERFOLGS-KRITERIEN CHECKLISTE

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Kategorie	Kriterium	Status
Erkennungsrate	$\geq 95\%$	[ ]
Erkennungsrate (Variationen)	$\geq 90\%$	[ ]
Falsch-Positive (ähnlich)	<5%	[ ]
Falsch-Positive (Konv.)	<1/Stunde	[ ]
Umgebungs-Robustheit	$\geq 85\%$	[ ]
Keine Doppel-Erkennungen	0 von 100	[ ]
Langzeit-Stabilität	$\geq 95\%$	[ ]
Latenz	<100ms	[ ]
CPU-Auslastung	<20%	[ ]
Memory-Auslastung	<200MB	[ ]

GESAMT: \_\_ /10 Kriterien erfüllt

PRODUKTIONSREIF: [ ] Ja (10/10) [ ] Nein (<10/10)

#### Empfohlene Kriterien (Optional)

Für exzellente Performance:

#### ERWEITERTE KRITERIEN (OPTIONAL)

Datum: \_\_\_\_\_ Tester: \_\_\_\_\_

Kategorie	Kriterium	Status
Erkennungsrate	$\geq 98\%$	[ ]
Falsch-Positive (Konv.)	<0.5/Stunde	[ ]
Latenz	<50ms	[ ]
CPU-Auslastung	<10%	[ ]
Memory-Auslastung	<100MB	[ ]
Funktioniert bis 4m Abstand	$\geq 80\%$	[ ]
Funktioniert in allen Räumen	$\geq 90\%$	[ ]

GESAMT: \_\_ /7 erweiterte Kriterien erfüllt

EXZELLENT: [ ] Ja ( $\geq 5/7$ ) [ ] Nein (<5/7)

# Test-Protokoll

## Vollständiges Test-Protokoll

### WAKE-WORD TEST-PROTOKOLL

=====

Projekt: Computer Voice Assistant

Wake-Word: "Computer"

Methode: [ ] Porcupine [ ] OpenWakeWord

### TESTER-INFORMATION

-----  
Name: \_\_\_\_\_

Datum: \_\_\_\_\_

Uhrzeit: \_\_\_\_\_

System: \_\_\_\_\_

Mikrofon: \_\_\_\_\_

### KONFIGURATION

-----  
Sensitivity: \_\_\_\_\_

Cooldown: \_\_\_\_\_ Sekunden

Sample Rate: \_\_\_\_\_ Hz

Threshold: \_\_\_\_\_ (falls OpenWakeWord)

### TEST-ERGEBNISSE

#### 1. Erkennungsrate-Tests

[ ] Test 1.1: Basis (\_\_\_%)

[ ] Test 1.2: Tonlagen (\_\_\_%)

[ ] Test 1.3: Geschwindigkeit (\_\_\_%)

[ ] Test 1.4: Abstände (\_\_\_%)

#### 2. Falsch-Positiv-Tests

[ ] Test 2.1: Ähnliche Wörter (\_\_\_%)

[ ] Test 2.2: Konversation (\_\_\_ /Stunde)

[ ] Test 2.3: Hintergrund (\_\_\_ /Stunde)

#### 3. Umgebungs-Tests

[ ] Test 3.1: Räume (\_\_\_%)

[ ] Test 3.2: Lautstärken (\_\_\_%)

#### 4. Stress-Tests

[ ] Test 4.1: Wiederholungen (\_\_\_%)

[ ] *Test 4.2: Langzeit* (\_\_\_%)

#### 5. Performance-Tests

[ ] Test 5.1: Latenz (\_\_\_ms)

[ ] *Test 5.2: Ressourcen* (\_\_\_% CPU, \_\_\_MB)

#### **GESAMTBEWERTUNG**

*Produktionsreif:* [ ] Ja [ ] Nein

*Empfehlung:* [ ] Deployment [ ] Weitere Optimierung

#### **NOTIZEN & BEOBACHTUNGEN**

---

---

---

---

#### **NÄCHSTE SCHRITTE**

---

---

---

Unterschrift: \_\_\_\_\_

---

## Dokumentende

---

Seite {page}