Abstract

Beschreibung und Anleitung für die Nutzung des Forensic WhisperDeskZH

ForMaLTeC

leo.engelberger@uzh.ch

Forensic WhisperDeskZH

User Manual

Contents

[1. Übersicht 2](#_Toc206518367)

[Wichtigste Merkmale: 2](#_Toc206518368)

[2. Voraussetzungen 2](#_Toc206518369)

[Minimale Systemanforderungen 2](#_Toc206518370)

[Empfohlene Systemanforderungen 2](#_Toc206518371)

[3. Wichtige Begriffe & Symbole 3](#_Toc206518372)

[4. Installation 4](#_Toc206518373)

[Schritt für Schritt 4](#_Toc206518374)

[5. Erste Schritte 4](#_Toc206518375)

[Erster Start 4](#_Toc206518376)

[Einfache Transkription 4](#_Toc206518377)

[6. User Interface 5](#_Toc206518378)

[Steuerelemente 5](#_Toc206518379)

[Zuhören Starten/Beenden 5](#_Toc206518380)

[Diktat Starten/Beenden 5](#_Toc206518381)

[Mikrofonauswahl 6](#_Toc206518382)

[Modelauswahl 6](#_Toc206518383)

[Sprachauswahl 6](#_Toc206518384)

[Reset-Knopf 7](#_Toc206518385)

[7. Sprachbefehle & Satzzeichen 7](#_Toc206518386)

[Sprachbefehle im Zuhörer Modus 7](#_Toc206518387)

[Sprachbefehle im Diktatmodus 7](#_Toc206518388)

[Standartbefehle 7](#_Toc206518389)

[Benutzerdefinierte Befehle 8](#_Toc206518390)

[Die keywords.xml Datei im Detail 8](#_Toc206518391)

[8. Fehlerbehebung 8](#_Toc206518392)

# Übersicht

Der Forensic WhisperDeskZH ist ein Add-In für Microsoft Word, das die Transkription von Sprache in Text in Echtzeit ermöglicht. Es verwendet fortschrittliche KI-Modelle (Whisper), um Ihre gesprochenen Worte direkt als Text in Word-Dokumenten einzufügen.

## Wichtigste Merkmale:

* **Echtzeit-Transkription** direkt in Word Dokumente
* **Mehrsprachige Unterstützung**
* **Verschiedene KI-Modelloptionen** für optimale Leistung

# Voraussetzungen

## Minimale Systemanforderungen

**Betriebssystem:** Windows 10 oder höher

**Microsoft Word:** 2016 oder höher

**Arbeitspeicherpeicher:** 8GB RAM

**Speicherplatz:** 8 GB freier Speicherplatz

**GPU:**  nicht notwendig (aber empfohlen)

**Audio:** Funktionierendes Mikrofon oder Audio-Eingabegerät

## Empfohlene Systemanforderungen

**Betriebssystem:** Windows 10 oder höher

**Microsoft Word:** 2016 oder höher

**Arbeitspeicherpeicher:** 16 GB RAM

**Speicherplatz:** 16 GB freier Speicherplatz

**GPU:** Nvidia GPU mit CUDA-Kernen und 8 GB GPU Speicher

**Audio:** Funktionierendes Mikrofon oder Audio-Eingabegerät

# Wichtige Begriffe & Symbole

**Diktatmodus:** Im Diktatmodus, transkribiert das Add-in sämtlcihe vom Mikrofon aufgenommene Sprache und fügt den transkribierten Text direkt in das geöffnete Word Dokument an der Stelle des Cursors ein.

**Zuhörmodus:**  Der Zuhörmodus bezeichnet den Modus, in welchem das Add-in bereits mithört, den transkribierten Text aber noch nicht in das geöffnete Word Dokument einfügt. Erst wenn das Add-in Durch den Sprachbefehl «Diktat Starten» in den Diktatmodus versetzt wird, wird Text im Dokument eingefügt.

**Minimal Audio Chunk Size:** Für die Echtzeit Transkription, muss die Audioaufnahme kurz unterbrochen und ausgelesen werden. Die Minimal Audio Chunk Size bestimmt die minimale Zeitspanne in Sekunden, welche vergehen muss, dass das Add-in eine Transkription startet. Kürzere Werte führen zu häufigerer und schneller Transkription ins Word Dokument, im Gegenzug sinkt aber die Genauigkeit. Mit grösseren Werten wird die Transkription weniger häufig aktualisiert, die Genauigkeit nimmt aber zu.

**Silence Threshold:** Um die Genauigkeit der Transkription zu erhöhen versucht das Add-in Pausen zwischen Wörtern und Sätzen zu erkennen. Dies soll verhindern, das Worte abgeschnitten und somit fehlerhaft transkribiert werden. Der Silence Threshold bestimmt, wie lange es Still sein muss, dass das Plugin von einer Pause zwischen Worten ausgeht.

# Installation

## Schritt für Schritt

1. Die neusten Versionen des Add-Ins finden sie unter: [**Releases · ForMaLTeC/Forensic-WhisperDeskZH-Word-Add-In**](https://github.com/ForMaLTeC/Forensic-WhisperDeskZH-Word-Add-In/releases)
2. **Wählen sie den Gewünschten Release** aus der Release Liste.
3. **Laden Sie ALLE Teile herunter** welche nach folgendem Muster bennant sinD: VSTO-Addin-Package-[Versions-Nummer].zip.001, .002 usw., der Source-Code wird für die Installation nicht benötigt.
4. **Speichern Sie ALLE Teile im selben Verzeichnis/Ordner** auf Ihrem Computer.
5. **Entpacken Sie die Dateien mit 7-Zip** – klicken Sie mit der rechten Maustaste auf die Datei .001 und entpacken Sie sie lokal auf ihren Rechner.
6. **Öffnen Sie den entpackten Ordner.**
7. **Führen Sie „setup.exe“\*\*** aus den extrahierten Dateien **als Administrator** aus.
8. **Folgen sie den Anweisungen des Installers**

**WICHTIG: die setup.exe Datei nicht aus dem Ordner herausziehen. Sie muss neben der .vsto Datei und dem ApplicationFiles Directory ausgeführt werden.**

Ist die Installation erfolgreich abgeschlossen sollt bei einem Neustart von Word nun die Menu Leiste „Add-Ins“ hinzugefügt worden sein. Ist diese nicht sichtbar ist die Installation fehlgeschlagen.

# Erste Schritte

## Erster Start

1. Stellen Sie sicher, dass ein Mikrofon an ihrem PC angeschlossen ist und funktioniert.

2. Öffnen Sie Microsoft Word wie gewohnt.

3. Klicken Sie auf die Registerkarte „Add-Ins“ in der Multifunktionsleiste

4. Warten Sie, bis das Add-In initialisiert ist (dies kann beim ersten Start einige Augenblicke dauern)

## Einfache Transkription

1. **Öffnen Sie ein Word-Dokument** oder erstellen Sie ein neues Dokument

2**. Positionieren Sie den Cursor** an der Stelle, an der der transkribierte Text erscheinen soll

3. **Wählen Sie Ihr Mikrofon** aus der Dropdown-Liste

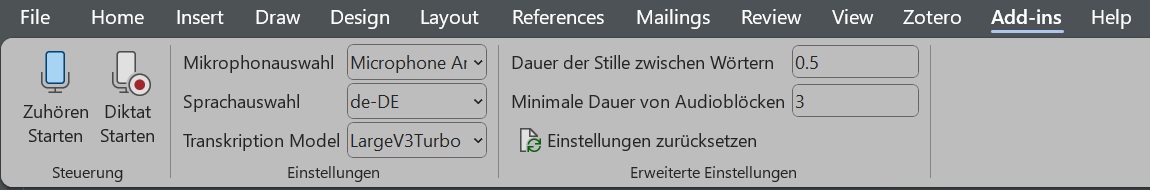
4. Klicken Sie auf **„Diktat starten“.**

5**. Beginnen Sie zu sprechen** - Ihre Worte werden im Dokument erscheinen

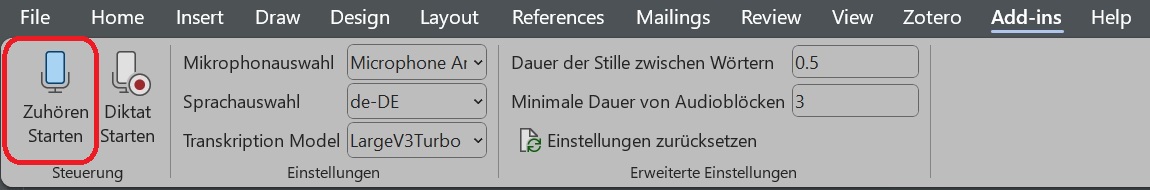
6. **Klick „Transkription stoppen“,** wenn Sie fertig sind

# User Interface

## Steuerelemente

Die Multifunktionsleiste des Add-ins enthält die folgenden Steuerelemente:

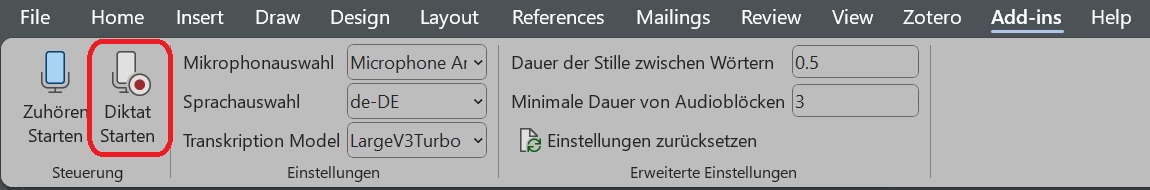
### Zuhören Starten/Beenden



Mit dem Knopf «Zuhören Starten» kann das Add-In in einen Zuhörer Modus versetzt werden. Befindet sich das Add-In in diesem Zuhörer Modus, transkribiert das Add-In sämtliche über das ausgewählte Mikrofon aufgezeichnete Sprache. Das Resultat der Transkription wird aber nicht in das aktive Word Dokument eingefügt. Erst wenn das Plugin den Sprachbefehl **«Diktat Starten»** wahrnimmt, beginnt das Einfügen des transkribierten Textes. Um das Einfügen zu beenden, muss der **Sprachbefehl «Diktat Beenden»** genutzt werden. Wird der Zuhörer Modus nicht beendet, so kann jederzeit wieder ein Diktat begonnen werden.

* **Zuhören Starten:** Der Zuhörmodus wird aktiviert.
* **Zuhören Beenden:** Der Zuhörmodus wird wieder beendet.
* Der Status ändert sich je nach aktuellem Transkriptionsstatus

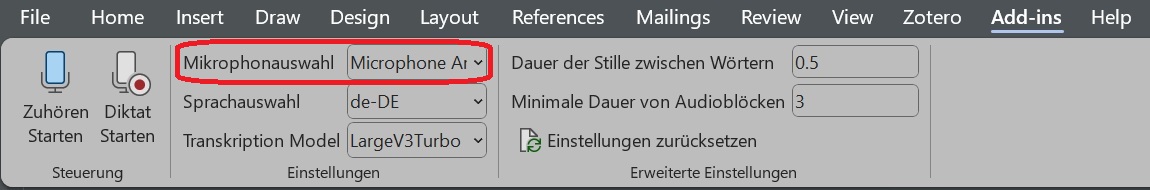
### Diktat Starten/Beenden



Mit dem Diktat Starten/Beenden Knopf lässt sich das AddIn direkt in den Diktatmodus versetzten. In diesem Diktat Modus wird sämtliche, über das ausgewählte Mikrofon aufgezeichnete Sprache transkribiert und die Transkription wird direkt beim Cursor in das Word Dokument eingefügt.

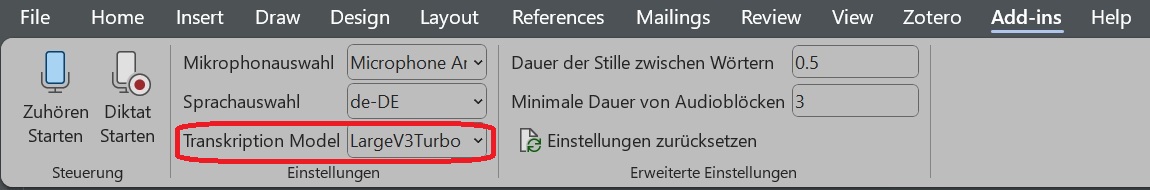
* **Diktieren starten:** Beginnt die Spracherkennung in Echtzeit.
* **Diktieren stoppen:** Beendet die Transkription und speichert die Einstellungen
* Der Status ändert sich je nach aktuellem Transkriptionsstatus

### Mikrofonauswahl



* Dropdown-Listfe der verfügbaren Audio-Eingabegeräte
* Wählen Sie Ihr bevorzugtes Mikrofon, bevor Sie mit der Transkription beginnen
* Änderungen werden sofort übernommen

### Modelauswahl

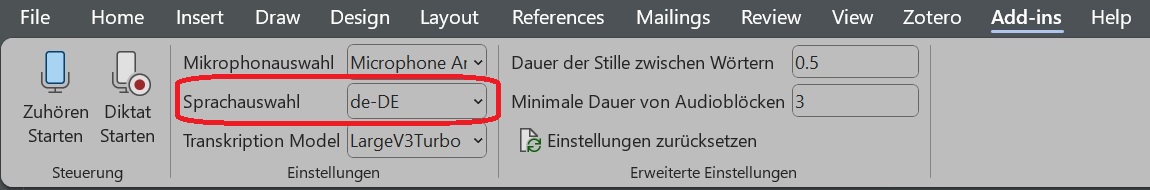


Wählen Sie je nach Bedarf aus verschiedenen KI-Modellen:

* **Tiny:** Sehr schnell, aber sehr ungenau.
* **Small**: Bessere Genauigkeit, moderate Geschwindigkeit.
* **Base**: Gutes Gleichgewicht zwischen Geschwindigkeit und Genauigkeit.
* **Medium**: Hohe Genauigkeit, langsamere Verarbeitung
* **Large** **V3**: Höchste Genauigkeit, langsamste Verarbeitung
* **Large** **V3** **Turbo**: Beste Genauigkeit bei verbesserter Geschwindigkeit

**WICHTIG: Die Modelle «Large V3» und «Large V3 Turbo» benötigen viele Ressourcen! Es sollten die empfohlenen Systemanforderungen erfüllt sein.**

### Sprachauswahl



Es besteht die Möglichkeit zwischen verschiedenen Sprachen auszuwählen.

Aktuell sind folgende Sprachen verfügbar:

* Deutsch
* Englisch
* Französisch
* Italienisch
* Spanisch

### Modeleinstellungen

Der Forensic WhisperDeskZH ermöglicht das anpassen zweier Einstellungen, welche direkten Einfluss auf die Echtzeittranskription nehmen.

#### Dauer der Stille zwischen Wörtern

Der Wert „Dauer der Stille zwischen Wörtern„ definiert die Anzahl Sekunden, welche vom Add-in als Pause zwischen Wörtern wahrgenommen werden. Dies hat direkten Einfluss darauf, wo im Satz das Add-in die Audioaufnahme transkribiert.

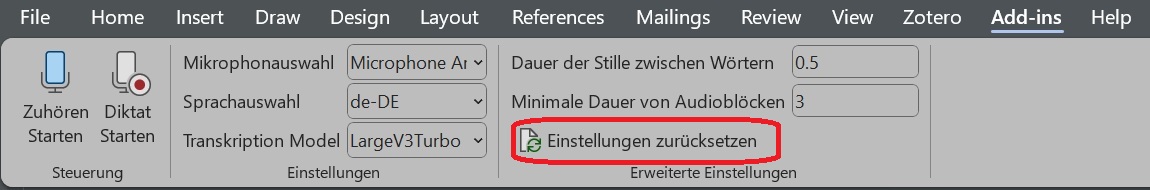
Für langsame Sprecher sollte dieser Wert erhöht werden, spricht man eher langsam kann er reduziert werden.

#### Minimale Dauer von Audioblöcken

Die Minimale Dauer von Audioblöcken ist die untere Grenze für die Häufigkeit in der Audioblöcke transkribiert werden. Ein tieferer Wert steigert die Kadenz in welcher neu Transkribierter Text eingefügt wird, ein hoher Wert senkt diese.

**ACHTUNG: Mit höherer Kadenz sinkt die Genauigkeit der Transkription!**

### Reset-Knopf



Mit dem Reset-Knopf lassen sich die erweiterten Einstellungen einfach zurücksetzten. Durch ein Drücken auf diesen Knopf werden folgende Einstellungen zurückgesetzt:

* Silence Threshold wird auf 0.5 Sekunden gesetzt.
* Minimal Audio Chunk Size wird auf 3 Sekunden gesetzt.

**ACHTUNG: Vorherige Einstellungen können nicht wieder hergestellt werden.**

# Sprachbefehle & Satzzeichen

## Sprachbefehle im Zuhörer Modus

* **Diktat Starten:** Das Add-in wechselt vom Zuhör- in den Diktatmodus
* **Diktat Beenden:** Das Add-in wechselt vom Diktatmodus in den Zuhörmodus

**ACHTUNG: Beide Befehle funktionieren nur wenn der Zuhörmodus über den «Zuhörmodus Starten/Beenden» Knopf aktiviert wurde.**

## Sprachbefehle im Diktatmodus

Das System erkennt Sprachbefehle für Interpunktion und Formatierung. Entdeckt das Add-in eines der folgenden Worte im transkribierten Text, so wird diese durch das entsprechende Satzzeichen ersetzt.

### Standartbefehle

* „Punkt“ → .
* „Komma“ → ,
* „Ausrufezeichen“ → !
* „Fragezeichen“ → ?
* „Doppelpunkt“ → :
* „Semikolon“ → ;
* „Bindestrich“ → -
* „Abstand“ → [Neuer Absatz]

### Benutzerdefinierte Befehle

Sie können benutzerdefinierte Sprachbefehle hinzufügen, indem Sie die Datei Keywords.xml ändern.

**WICHTIG: Das Add-in kann nur ersetzten. Sprachbefehle welche Word selbst kontrollieren (z.B. «Dokument speichern») sind aktuell nicht möglich.**