

# Text-to-Speech

Sprachsynthese ist die künstliche Generierung von Sprache mit Hilfe eines Rechners. Da mit Hilfe dieser Technologien ganze Texte vorgelesen werden können, nennt man diesen Prozess auch Text-to-Speech (TTS). Moderne TTS-Systeme sind sehr leistungsfähig; die gut verständlichen und natürlich klingenden Stimmen haben mit den blechernen Sprechetönen früherer Heimcomputer nicht mehr viel gemein.

## Anwendungsbeispiele

Sprachsynthese kann einen wichtigen Beitrag zur barrierefreien Nutzung des Rechners leisten, z. B. wenn beliebige Webseiten oder Programmmenüs für Menschen mit Sehbehinderung vorgelesen werden. Die Sprachsynthese hat auch in viele kommerzielle Anwendungen Einzug erhalten, z. B. dem Vorlesen von SMS-Nachrichten, Abrufen von E-Mails, Wetter- oder Börsendaten per Mobiltelefon oder dem Vorlesen von Fahrhinweisen durch Navigationssysteme.

## Sprachsynthese für E-Learning



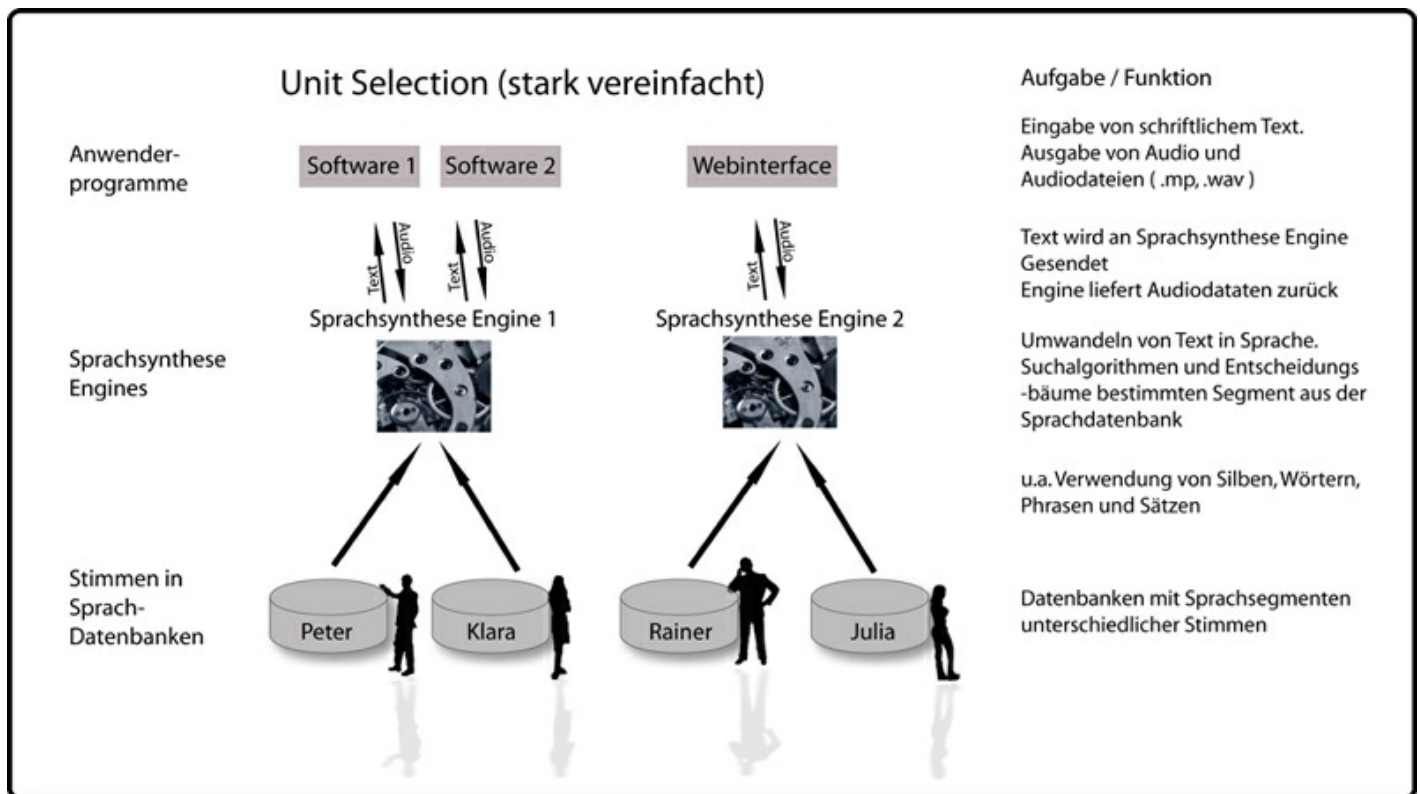
Im E-Learning kann Text-To-Speech vielseitig eingesetzt werden, zum Beispiel um schriftliche Unterlagen wie Vorlesungsskripte in Tondateien umzuwandeln. Die Inhalte können per Podcast bereitgestellt und auf mobilen Wiedergabegeräten abgespielt werden. Für das Lernen von Fremdsprachen sind vom Computer vorgelesene Wörter und Sätze ebenfalls geeignet wenn die generierte Sprache exakt ist. Für die Produktion multimedialer Lernangebote, z. B. Animationen, interaktive Darstellungen oder Software-Demos, sind Computersprecher außerdem eine kostengünstige und unkomplizierte Alternative zur Sprachaufnahme im Tonstudio. Dies gilt auch für die Vertonung von Avataren, also kleinen digitalen (Comic-)Helfern, die

durch ein Lernprogramm führen können. Leider erlauben die Lizenzbedingungen vieler Text-To-Speech-Programme keine kommerzielle Verwendung der generierten Audiodateien. Für die Produktion von Prototypen und den privaten Gebrauch eignen sie sich aber hervorragend.

## Qualität der Sprachsynthese

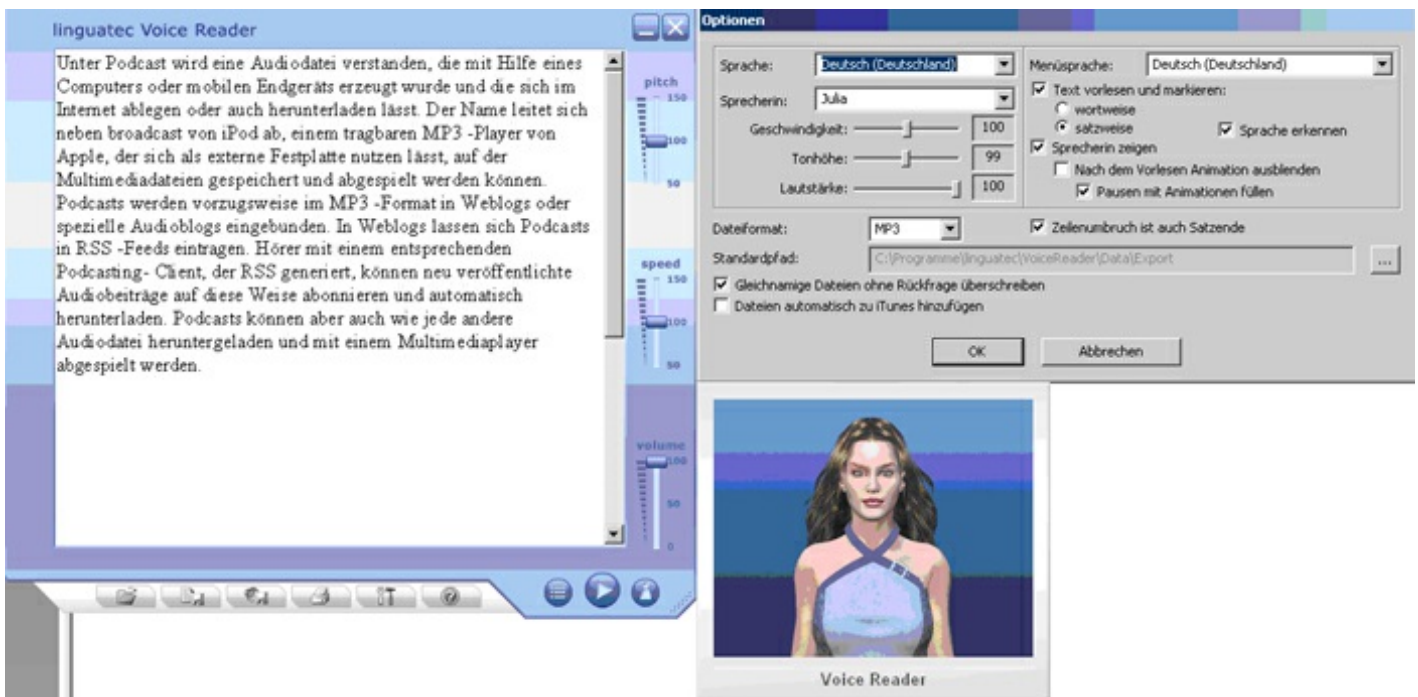
Die Qualität der generierten Sprache variiert sehr stark zwischen den unterschiedlichen Systemen. Die wichtigsten Qualitätskriterien sind die Natürlichkeit und die Verständlichkeit der gesprochenen Sprache. Ursachen für die Qualitätsunterschiede sind zum einen die verschiedenen Verfahren, mit denen die Sprache erzeugt wird, zum anderen die verwendete Datenbasis. Bei einigen Verfahren werden z. B. häufig gesprochene Wörter nicht künstlich generiert sondern liegen als eine Spracheinheit bereits digitalisiert vor. Je mehr Wörter verfügbar sind, desto wahrscheinlicher wird eine natürliche, fehlerfreie Aussprache. Je nach Verfahren wird entweder eher eine Natürlichkeit oder gute Verständlichkeit der Sprache erreicht. Weitere Qualitätsfaktoren sind eine angenehme

Stimmfrage, gut getroffene Sprachmelodie sowie die korrekte Aussprache und Umwandlung von Abkürzungen, Satzzeichen oder Zahlen.



## Sprachsynthese-Software und -Engines

Für die eigentliche Erzeugung der gesprochenen Worte kommen Sprachsynthese-Engines zum Einsatz. Bei diesen Engines handelt es sich um Softwaremodule, die von verschiedenen Herstellern angeboten werden. Für die meisten Engines sind mehrere weibliche und männliche Stimmen verfügbar. Die künstlichen Stimmen haben menschliche Namen, zum Beispiel Klara und Reiner (von AT&T <http://www.naturalvoices.att.com/>) oder Anna und Stefan (Nuance <http://www.nuance.com/>). Die Sprachsynthese Engines können Sie nicht direkt verwenden. Stattdessen gibt es verschiedene Softwarepakete, die den Einsatz von Sprachsynthese Engines erlauben. Die Programme bieten in der Regel eine Benutzerschnittstelle, in der die zu sprechenden Texte eingegeben werden können. Die Software schickt diesen Text dann an eine der verfügbaren Sprachsynthese-Engines. Während es sehr viele unterschiedliche Softwarepakete für Sprachsynthese gibt, ist die Anzahl der Engines geringer. Die meisten Softwareprodukte greifen auf die gleichen Engines zurück. Die Programme unterscheiden sich daher nur in der Benutzerfreundlichkeit, jedoch nicht in der Sprachqualität, da diese allein von den Engines abhängt. Beim Kauf einer Software sollte man darauf achten, dass eine hochwertige Engine mit zum Lieferumfang gehört. Der Linguattec Voice Reader (Produktsteckbrief) verwendet zum Beispiel die gut klingende Sprachsynthese von SVOX (SVOX <http://www.svox.ch/>).



Viele der verfügbaren Stimmen können Sie kostenlos im Internet testen. Auf den Seiten der Hersteller gibt es in der Regel ein Formular, in das Sie eigene Kurztexte eingeben und eine Stimme wählen können. Der Text wird dann an die jeweilige Sprachsynthese-Engine gesendet. Wenn Ihnen die Sprachqualität einer Stimme gefällt, dann müssen Sie beim Softwarekauf nur darauf achten, dass diese spezielle Stimme (z. B. „Dagmar“) mit zum Lieferumfang gehört. Während für deutschsprachige Stimmen auch deutsche Rufnamen verwendet werden, erkennt man englischsprachige Stimmen meistens an Namen wie „Paul“ oder „Mary“. Sowohl die Engines als auch die Stimmen sind für verschiedene Sprachen optimiert, wobei die Besonderheiten der jeweiligen Aussprache berücksichtigt werden. Einige Hersteller berücksichtigen auch die Unterschiede zwischen britischer und amerikanischer Aussprache.

## Weitere Informationen

- Voice Reader – Produktsteckbrief
- Umfangreiche Auswahl an Sprachbeispielen:  
<http://ttssamples.syntheticspeech.de/deutsch/index.html>
- Gute Sprachengines:  
 RealSpeak  
 Nuance Vocalizer
- Open Source Engines:  
 festvox  
 MARY Text-to-Speech

Letzte Änderung: 15.04.2015

## Zitation

e-teaching.org Redaktion (2015). Text-to-Speech. Zuletzt geändert am 15.04.2015. Leibniz-Institut für Wissensmedien:  
[https://www.e-teaching.org/technik/aufbereitung/text/Text-to-Speech/index\\_html](https://www.e-teaching.org/technik/aufbereitung/text/Text-to-Speech/index_html). Zugriff am 16.07.2018