HIS ▸ HIJP ▸ AIGP

Harmonisierung der Informatik in der Strafjustiz
Harmonisation de l'informatique dans la justice pénale
Armonizzazione dell'informatica nella giustizia penale

softfakt
Potenzial effektiv entfalten

**Softfakt GmbH**
Rosenbergstrasse 75
8498 Gibswil

+41 55 245 11 66

info@softfakt.ch
www.softfakt.ch

# Speech-to-Text in der Justiz

Rapport

**Auteur:** Ralph Wildhaber
**Version**: 1.0
**Date:** 31.08.2021

## Sommaire



## Tableaux

## Figures

# 1 Bases du Speech-to-Text (reconnaissance vocale et transcription)

Dans bien des domaines, le Speech-to-Text est de plus en plus apprécié et important. Le sujet est souvent lié aux activités de numérisation ou à l'optimisation des processus dans les entreprises. Dans le quotidien du secteur privé également, les systèmes de reconnaissance vocale sont désormais omniprésents. L'offre disponible sur le marché est donc d'autant plus diversifiée.

## 1.1 Terminologie et fonctionnement

Les termes «Speech-to-Text», «Speech Recognition», «Voice Recognition», reconnaissance automatique de la parole, transcription audio ou système de transcription désignent au fond la même chose: la transcription assistée par ordinateur de la voix humaine.

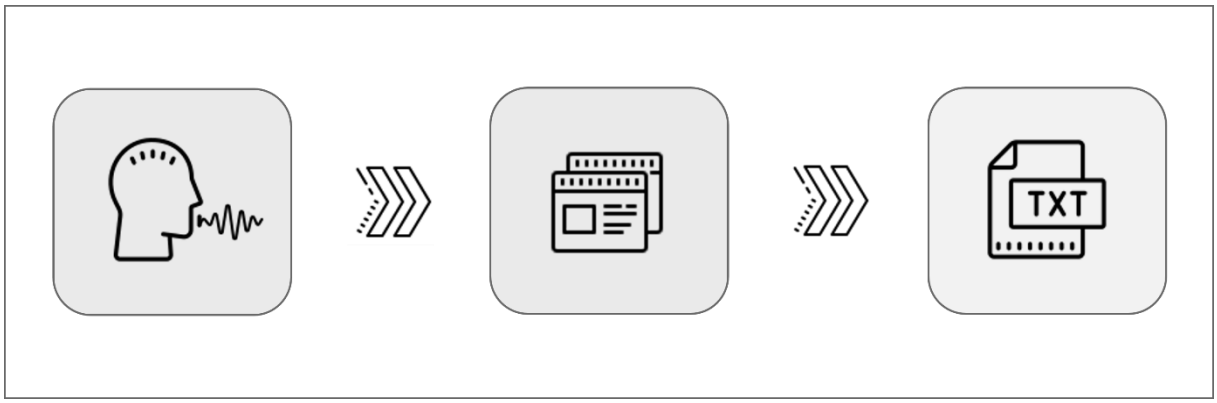


**Figure 1: déroulement de la reconnaissance vocale automatique**

A cette occasion, un signal audio (parlé) est capté au moyen d'un ordinateur et traité par un logiciel de reconnaissance de la parole pour produire un texte lisible. De tels programmes reposent sur des procédures à plusieurs niveaux qui ont recours à divers algorithmes. L'analyse passe en règle générale par les trois modèles suivants l'un après l'autre:

- **Modèle acoustique** pour décomposer le signal audio
- **Lexique** pour identifier les mots
- **Modèle linguistique** pour construire une phrase sur la base d'une suite de trois mots

Bien que l'informatique étudie depuis des décennies déjà la reconnaissance vocale, l'importance de celle-ci, de même que sa diffusion, n'ont cessé de prendre de l'ampleur ces dernières années. Les raisons de cette expansion résident d'une part dans les performances accrues des ordinateurs, qui sont en mesure d'effectuer les analyses susmentionnées en une fraction de seconde. D'autre part, la diffusion des technologies mobiles (smartphone, tablette) a contribué à l'accélération du développement du «Speech-to-Text».

Selon la solution retenue, la transcription a lieu directement lors de la dictée (en simultané) ou est effectuée quelques minutes après avoir chargé le fichier audio correspondant.

## 1.2 Délimitations de la reconnaissance vocale

Les systèmes Speech-to-Text doivent clairement être distingués des procédures biométriques servant à l'identification (reconnaissance du locuteur) ou des analyses vocales. De telles solutions de reconnaissance vocale ne sont pas non plus (encore) utilisées pour la traduction.

## 1.3 Composants d'infrastructure

Si l'on considère ce dont un utilisateur d'un logiciel de reconnaissance vocale a besoin en termes d'infrastructure, la situation est des plus simples. Selon le type d'installation, il y a de petites différences.

### 1.3.1 Matériel

En termes de matériel, il faut un microphone en plus d'un ordinateur relativement récent. Il est recommandé de recourir ici à un micro-casque. Ces appareils sont optimisés pour l'enregistrement vocal et moins susceptibles d'être perturbés par des bruits environnants en raison de leur positionnement. Mentionnons toutefois ici que d'excellents résultats de transcription peuvent aussi être obtenus grâce aux microphones intégrés aux ordinateurs (portables) ou aux webcams.

Si l'on considère la situation en termes d'exploitation, les choses sont un peu différentes. De nombreuses solutions, justement dans le domaine des logiciels locaux de dictée, permettent de se baser sur un serveur dans le centre de calcul propre. Les exigences envers un dispositif correspondant satisfont aux standards usuels. Il est donc possible de les exploiter sans problème dans des environnements virtuels.

### 1.3.2 Logiciels

En fonction du domaine d'application et du type de services, des logiciels différents sont nécessaires pour l'enregistrement ou la transcription de l'information vocale. On fait la distinction entre des solutions web et des programmes à installer localement (applications client).

Le grand avantage des solutions web est que les programmes peuvent être utilisés avec des navigateurs web courants et sont donc indépendants du système d'exploitation.

Par ailleurs, certaines solutions sur le marché nécessitent l'installation du logiciel de reconnaissance vocale sur l'ordinateur cible. Il convient de noter que les

différents produits ne sont pas tous disponibles dans des versions compatibles avec les systèmes d'exploitation classiques (Windows, MacOS, Linux).

## 1.4 Développements et technologies

En matière de reconnaissance vocale, les développements sont en cours depuis de nombreuses décennies déjà. Ce n'est que grâce aux meilleures performances des ordinateurs modernes qu'il a été possible d'intégrer de nouvelles méthodes de transcription vocale. Cette étape de développement est aussi la raison pour laquelle la reconnaissance vocale automatique a gagné en importance et va continuer de le faire. Si, précédemment, il était surtout fait appel à des modèles statiques, les méthodes appliquées aujourd'hui sont plus gourmandes en puissance de calcul car basées sur l'intelligence artificielle.

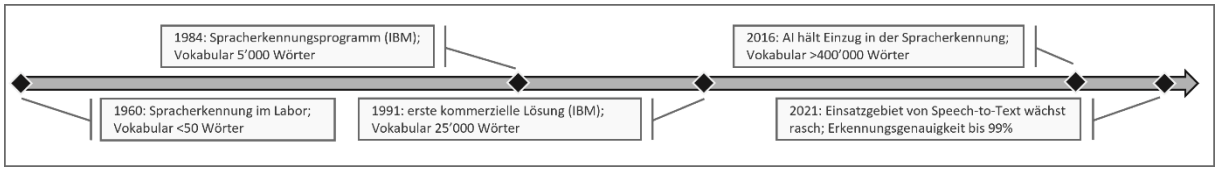


**Figure 2: Progrès de la reconnaissance vocale automatique**

Cela a un net avantage. Les systèmes basés sur l'IA sont disponibles pour l'utilisateur sans phase d'apprentissage de la part du logiciel. Précédemment, le locuteur et l'ordinateur devaient «apprendre à se connaître» au moyen d'un module d'entraînement. En d'autres termes, des mots et phrases prédéfinis devaient être prononcés pour que le système comprennent comment l'orateur s'exprimait. Une autre limitation en résultait. Le logiciel ne pouvait fournir de bons résultats que pour le locuteur qu'il connaissait. Si un autre utilisateur avait recours à la même configuration (sans entraînement), les résultats de la transcription étaient nettement moins bons, car le système n'était pas préparé à ce locuteur.

Hormis la suppression de l'entraînement initial pour configurer le système, les solutions modernes présentent un autre avantage. Elles apprennent en permanence comment l'utilisateur s'exprime (mots et tournures préférés). Chaque transcription est intégrée au profil linguistique personnel d'un locuteur et peut ensuite être réutilisée. Le taux de reconnaissance s'améliore ainsi en continu. Plusieurs solutions atteignent aujourd'hui déjà des taux de reconnaissance pouvant atteindre 99%. Cette valeur dépend évidemment de plusieurs facteurs: sujet, longueur, complexité du contenu, etc.

## 1.5 Gamme de fonctions

Reconnaissance vocale n'est pas synonyme de reconnaissance vocale. Techniquement, cette affirmation tend à être fausse. Des concepts identiques sont en effet de plus en plus mis en œuvre. Il en va différemment en ce qui concerne les domaines d'application. Les différences concernent différents

aspects, comme l'utilisation, l'étendue fonctionnelle, la complexité et le vocabulaire.

### 1.5.1 Fonction de base

En principe, il s'agit toujours **de convertir la langue orale en texte lisible**. L'utilisation conforme de tels systèmes permet aujourd'hui d'obtenir d'excellents résultats en termes de transcription. Cela signifie que les assistants vocaux (cf. ch. 2.1.1) fournissent des réponses adéquates à des questions simples. En essayant par exemple d'enregistrer un article médical spécialisé avec un assistant vocal comme Siri, le résultat obtenu demanderait beaucoup de travail de correction ultérieur.

### 1.5.2 Fonctions supplémentaires

Une analyse des solutions Speech-to-Text disponibles sur le marché indique que la gamme de fonctionnalités ne cesse de s'étoffer. Les anciennes caractéristiques de différenciation de certains produits se retrouvent de plus en plus souvent dans un grand nombre de produits. Il y a toutefois diverses fonctions utiles, qui ne sont d'une part pas disponibles dans toutes les solutions et, d'autre part, qui pourraient jouer un rôle déterminant dans l'évaluation d'un produit. La liste ci-après indique certaines de ces caractéristiques:

- Intégration de vocabulaires spécialisés (par ex. médecine, justice)
- Possibilité de compléter les vocabulaires existants (individualisation)
- Création d'un ou de plusieurs profils d'utilisateurs (par ex. pour une utilisation avec des langues différentes)
- Traitement de fichiers audio (mp3, wav, etc.)
- Traitement de fichiers vidéo (par ex. pour générer des sous-titres)
- Divers formats de sortie (rft, txt, doc, docx, etc.)
- Reconnaissance automatique des signes de ponctuation et des caractères de formatage
- Ordres vocaux individuels pour des modules de textes et signatures propres
- Ordres de formatage individuels
- Pilotage vocal du programme

Cette liste n'est nullement exhaustive. Il s'agit d'indiquer que la recherche de solutions Speech-to-Text adéquates, il est possible de prendre en compte un grand nombre de paramètres d'exigences.

# 2 Tour d'horizon du marché

## 2.1 Aperçu des produits

La répartition des systèmes de reconnaissance vocale peut grosso modo être divisée en trois domaines. Notons toutefois qu'avec cette catégorisation, il n'est pas possible d'effectuer une délimitation précise. Désormais, il y a des produits qui apparaissent et sont utilisés dans plusieurs catégories.

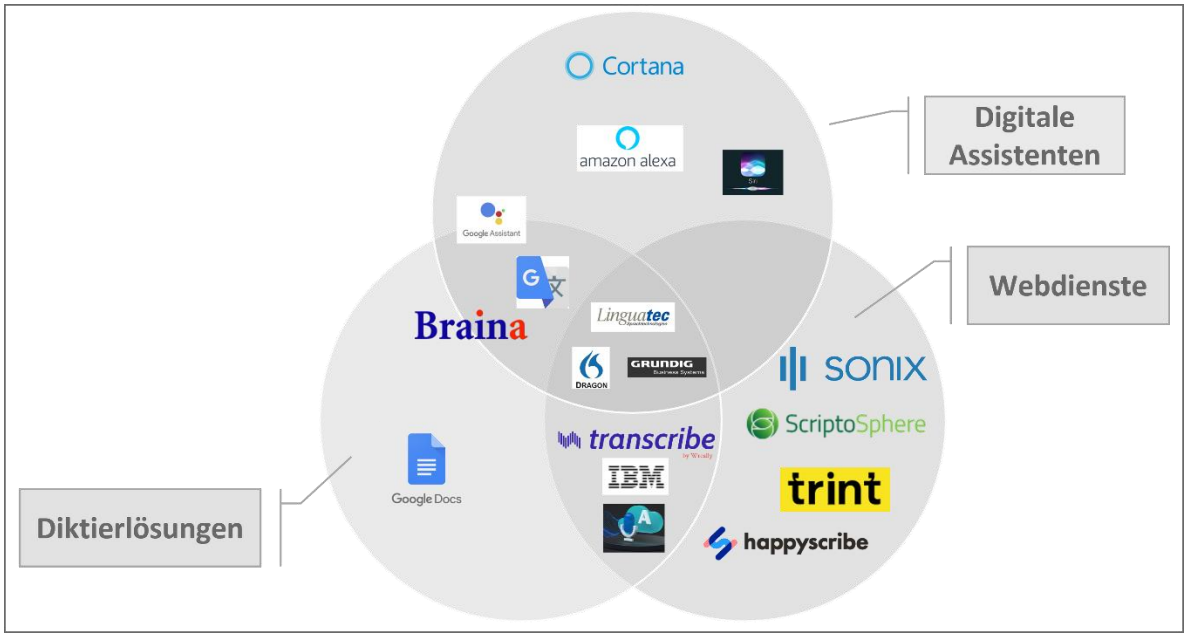


**Figure 3: Aperçu des produits en matière de reconnaissance vocale**

### 2.1.1 Assistant numérique

La catégorie des assistants numériques regroupe par ex. Siri d'Apple ou Alexa d'Amazon, etc., mais aussi les systèmes de navigation guidés par la voix dans les voitures. Ces dispositifs sont paramétrés pour exécuter des consignes très simples. Tant qu'il s'agit de lire un bulletin météo, de proposer une liaison ferroviaire, d'enregistrer un rendez-vous dans l'agenda ou de composer un numéro de téléphone, ces services donnent de très bons résultats. Ils sont certainement moins, voire pas du tout, adaptés pour l'enregistrement de textes plus longs ou la production de documents formatés correctement.

De plus, il n'est en règle générale pas possible d'étoffer le vocabulaire (quantité d'ordres vocaux) d'un tel assistant, contrairement à ce qui est le cas pour les deux autres catégories.

### 2.1.2 Service web

Les services web sont utiles pour transcrire des exposés, présentations, interviews ou des choses similaires. L'accent est clairement placé sur l'archivage par écrit de l'oral. Des exigences telles qu'un formatage propre ou la production

de textes orthographiés correctement sont moins importantes et ne sont pour l'heure remplies que par peu de services.

Un autre critère souvent cité en faveur de l'utilisation de solutions Speech-to-Text en ligne est leur disponibilité en tout lieu.

De nos jours, certains services web permettent d'étoffer le vocabulaire. La mise à disposition de collections de termes et expressions spécifiques fait par contre défaut.

### 2.1.3 Solutions de dictée

Dans les métiers médicaux et judiciaires, mais aussi dans d'autres domaines administratifs, des systèmes de dictée sont actuellement utilisés à large échelle. Pour les produits correspondants, on exige une reconnaissance textuelle élevée, y compris lors de l'utilisation de vocabulaires spécifiques. De plus, l'on s'attend à ce que les textes puissent être formatés par des instructions vocales.

En ce qui concerne l'individualisation du système, ces solutions offrent des avantages importants. En règle générale, il est possible d'étendre le vocabulaire en y ajoutant des expressions propres. De plus, il est possible de définir de nouvelles instructions vocales, qui peuvent être utilisées pour appliquer des formatages préférés ou intégrer des textes standard personnalisés.

Ces systèmes apprennent à connaître et transcrire les habitudes langagières du locuteur en recourant des méthodes d'apprentissage automatique (machine learning).

## 2.2 Comparaison des produits

| Produit | happyscribe | Trint | Transcribe | Dragon | Watson | Azure | Braina |
|---|---|---|---|---|---|---|---|
| Fabricant | Happy Scribe | Trint | Wreally | Nuance | IBM | Microsoft | Brainasoft |
| URL | happyscribe.com | trint.com | transcribe.wreally.com | nuance.com | ibm.com/cloud/watson-speech-to-text | azure.microsoft.com | brainasoft.com |
| Exigence système | Navigateur web | Navigateur web | Navigateur web | Application Client | Navigateur web | Navigateur web | Application Client |
| **Transcription audio** | | | | | | | |
| Langues | DE, FR, IT, EN (>60 langues) | DE, FR, IT, EN (env. 50 langues) | DE, FR, IT, EN (env. 70 langues) | DE, FR, IT, EN (env. 8 langues) | DE, FR, IT, EN (env. 10 langues; plusieurs dialectes) | DE, FR, IT, EN (env. 40 langues) | DE, FR, IT, EN (env. 100 langues) |
| Choix de la langue | Manuel | Manuel | Manuel / automatique | Manuel | Manuel | Manuel | Manuel |
| Vocabulaire spécialisé | non | non | non | Oui, justice et médecine | non | non | non |
| Vocabulaire extensible | Oui, max. 100 mots | Oui, max. 100 mots | non | oui | oui | oui | oui |
| Transcription simultanée | non | Oui (par streaming) | Oui (éditeur propre) | Oui (éditeur propre ou directement dans l'application cible) | oui | oui | non |
| Téléchargement de fichier | oui | Oui | oui | oui | oui | oui | non |
| Formats d'entrée (audio) | AAC, AIFF, FLAC, M4A, MP3, Ogg Vorbis, WAV, WMA, entre autres | AAC, M4A, MP3, WAV, WMA | AAC, AIFF, FLAC, M4A, MP3, Ogg Vorbis, WAV, WMA, entre autres | MP3, WAV, M4A, WMA, DSS, DS2, AIFF, M4V | MP3, MPEG, WAV, FLAC, OPUS | MP3, OPUS/OGG, FLAC, ALAW, MULAW | - |
| Formats de sortie | TXT, DOC, PDF, JSON, STL, SRT, VTT, entre autres | DOCX, SRT, VTT, TXT, STL, EDL, HTML, XML, CSV | DOC | RTF | Texte | Texte | Directement dans l'application cible Human Language Interface (Braina natif) |
| Formatage de texte | oui | oui | non | oui | En partie | non | non |
| Autres fonctions | | Traduction autom. | Approches de reconnaissance de plusieurs orateurs (mode expérimental) | Traitement par lot de plusieurs enregistrements audio | Approches de reconnaissance de plusieurs locuteurs (en anglais seulement) | Reconnaissance autom. des signes de ponctuation | |
| **Possibilités d'exploitation** | | | | | | | |
| SaaS / Cloud | oui | oui | oui | non | oui | oui | non |
| On Premise | non | non | non | non | oui | oui | non |
| Installation Client | - | - | - | oui | non | non | oui |
| **Prix / facturation** | | | | | | | |
| Licence | - | - | - | Home: env. CHF 220<br>Professional: env. CHF 750 | - | - | env. $ 300 (lifetime) |
| Utilisation | € 12 par heure | dès € 55 par mois | $ 20 par année et<br>$ 6 par heure | - | env. CHF 1.15 par heure | Heure: CHF 0.99 par heure<br>Cust: CHF 1.4 par heure | env. $ 60 par année (alternative) |

**Tableau 1: Comparaison de solutions de reconnaissance vocale sélectionnées**

### 2.2.1 Grundig Business Systems

Grundig Business Systems a une approche intéressante. Cette entreprise spécialisée dans la reconnaissance vocale propose une gamme de services qui repose sur le produit Dragon, de l'entreprise Nuance, qui est très répandu et leader du secteur. Sur le plan fonctionnel, la solution est identique à Dragon. En termes de configuration, le fournisseur met par contre des configurations optimisées à disposition. Il y a par exemple des vocabulaires adaptés pour une utilisation en Suisse alémanique (ss au lieu de ß).

### 2.2.2 Délimitations

Le Tableau 1 compare une sélection de produits des catégories Services web et Solutions de dictée. Les assistants numériques purs n'ont délibérément pas fait l'objet d'une analyse approfondie, car ils ne correspondent pas au domaine d'application pertinent en l'espèce.

# 3 Considérations concernant l'exploitation

Le paragraphe sur les aspects concernant l'exploitation se focalise sur les modèles d'exploitation et de coûts. Par ailleurs, certains facteurs sont mentionnés qui ont une incidence positive sur la qualité de la reconnaissance vocale.

## 3.1 Modèles d'exploitation

La gamme complète des possibilités d'exploitation envisageables pour les systèmes de reconnaissance vocale est aujourd'hui couverte. Tout est disponible sur le marché: de l'installation locale sur des ordinateurs fixes jusqu'aux applications mobiles en passant par la distribution par des environnements Citrix ou des services en cloud, exploités en régie propre ou par des tiers.

Les services web ne peuvent en règle générale pas être exploités dans des centres de calcul propres. Pour l'heure, les noms plus connus sur le marché sont à cet égard encore l'exception. En effet, tant IBM que Microsoft proposent une exploitation «on premise» (sur site) de leurs solutions de reconnaissance vocale.

En ce qui concerne les solutions de dictée, qui, outre une installation locale, proposent également un serveur, il est en règle générale possible d'intégrer cette infrastructure dans un centre de calcul propre. L'avantage d'une telle solution est certainement lié au fait que les exigences en matière de protection des données peuvent être respectées. De plus, une exploitation en régie propre permet une gestion indépendante du fournisseur des configurations du système et des utilisateurs ainsi que des droits d'accès. Habituellement, de telles solutions sont judicieuses pour enregistrer sur le serveur des vocabulaires communs et une configuration centrale.

## 3.2 Modèles de licence et coûts

Dans les deux catégories considérées de manière plus détaillée des solutions Speech-to-Text, différents modèles de licences et plans tarifaires sont appliqués. A cet égard, les différences de prix sont très importantes (cf. Tableau 1, paragraphe «Prix / facturation»).

### 3.2.1 Modèles de licence

En ce qui concerne les services web exploités à l'externe, les modèles de licence sont assez similaires. Habituellement, la facturation dépend de la durée de texte dicté (facteur temporel). Il arrive que des fournisseurs facturent une taxe de base ou annuelle (par ex. transcribe de la société Wereally).

En ce qui concerne les solutions exploitées en régie propre, des modèles de licence habituels sont appliqués. Les licences uniques (lifelong) sont répandues à large échelle. Ces plans tarifaires sont souvent complétés par des licences de

groupe ou annuelles. Dans cette configuration d'exploitation, il est par ailleurs habituel de conclure des contrats d'assistance et de maintenance.

### 3.2.2 Coûts

Compte tenu des modèles de facturation et des coûts plutôt bas qui en résultent par unité temporelle, les services web conviennent parfaitement pour essayer et découvrir les solutions Speech-to-Text. Proportionnellement, les frais pour une utilisation occasionnelle de tels services sont en outre faibles, en comparaison avec des solutions avec licences uniques ou annuelles. De tels produits en valent la peine en cas d'utilisation fréquente (plusieurs heures par jour). Pour être juste, il convient de relever qu'il s'agit ici de systèmes de reconnaissance vocale qui proposent des fonctions à plus large échelle. Typiquement, elles se retrouvent dans la catégorie des solutions de dictée.

Le Tableau 1 met en évidence la vaste gamme de coûts qui résulte de la comparaison de quelques produits seulement.

## 3.3 Qualité de la transcription

Les résultats obtenus par les solutions Speech-to-Text, qui sont déjà en grande partie étonnamment bons, peuvent encore être améliorés en suivant quelques conseils simples:

- Une élocution claire et distincte permet d'obtenir de bons résultats. Il est recommandé de s'exercer à dicter.
- Dicter avec un ton naturel et à une vitesse normale. Ni une diction particulièrement lente, ni un volume plus élevé n'améliorent les résultats; au contraire, ils entraînent plutôt une baisse de la qualité, car les modèles d'analyse sont fondés sur une diction naturelle.
- Les signes de ponctuation et les caractères de formatage doivent toujours être dictés.
- Il est recommandé d'apprendre (par cœur) les principaux ordres de formatage et raccourcis clavier.
- Un environnement silencieux ou constant permet d'obtenir de meilleurs résultats. Comme mentionné précédemment au ch. 1.3.1, un casque avec microphone peut améliorer significativement les résultats dans un environnement qui n'est pas tout à fait calme.

En complément aux points ci-dessus, voici quelques enseignements tirés d'une phase de test de plusieurs mois:

- De nombreux produits sont conçus de manière à pouvoir être développés et ajustés par l'utilisateur. Cette possibilité devrait impérativement être utilisée en ce qui concerne les vocabulaires.
- Des optimisations des commandes, raccourcis clavier, blocs de texte, etc. disponibles facilitent quelque peu le travail.

- Il vaut la peine de prendre du temps pour se familiariser avec l'outil.

# 4 Possibilités d'utilisation dans le domaine judiciaire

## 4.1 Travail de chancellerie et administration

Dans le domaine judiciaire, les solutions Speech-to-Text (solutions de dictée) peuvent très bien servir à remplacer les dictaphones souvent utilisés actuellement. Elles permettent par exemple à un juge de dicter et de produire une correspondance spécifique à un cas, des décisions, ordonnances et d'autres documents similaires. Il n'est dès lors plus nécessaire de procéder comme à l'accoutumée, à savoir la saisie par des employés du secrétariat des textes dictés.

Il va de soi qu'un système de reconnaissance vocale peut aussi être intégré au travail de secrétariat quotidien. La possibilité de faire appel à des textes prédéfinis par commande vocale permet de produire des courriers en quelques secondes. Il est possible de procéder de la même manière pour la rédaction d'e-mails.

## 4.2 Utilisation lors d'auditions et de procédures judiciaires

Dans le domaine judiciaire en particulier, certains cas concrets semblent parfaits pour le recours à la reconnaissance automatique de la parole. Il semble ainsi possible de simplifier la verbalisation d'auditions ou d'audiences au tribunal. D'autres exigences étant liées à ces cas concrets, il convient d'examiner si les systèmes correspondants dans leur version actuelle satisfont à celles-ci.

Dans les deux cas concrets mentionnés, il convient de tenir compte de deux éléments:

1. **Langue du locuteur / dialecte**: Que ce soit au cours des audiences ou lors des dépositions, il est fréquent de recourir non pas à un langage standard (langue écrite) mais à un dialecte personnel. Or, ces dialectes peuvent parfois fortement diverger sur le plan régional. En ce qui concerne le système, la diversité linguistique est ainsi encore plus élevée et la reconnaissance de nuances dialectales est un peu plus compliquée.
2. **Reconnaissance de plusieurs locuteurs**: Les systèmes de reconnaissance vocale ont aujourd'hui souvent recours à des profils d'utilisateurs et sont donc paramétrés pour un locuteur. Certes, un système peut reconnaître et transcrire les éléments prononcés par différentes personnes, mais une attribution fiable à tel ou tel locuteur (si une telle fonction est même proposée) n'est pas encore possible avec le degré de confiance souhaité.

Les deux points abordés ci-dessus sont connus des fournisseurs de systèmes de reconnaissance vocale. Ceux-ci investissent donc dans la recherche et le développement afin de mieux identifier les locuteurs et les dialectes. Grundig Business Systems collabore à cet égard avec le Fraunhofer-Institut.

Actuellement, de telles solutions permettant la reconnaissance automatique de plusieurs locuteurs et susceptible d'être utilisées dans la pratique ne sont toutefois pas encore disponibles sur le marché. IBM Watson semble assez avancé sur ce point, du moins pour les dialogues en anglais. En raison de leur diffusion restreinte, les langues minoritaires, comme le suisse allemand et ses dialectes, se tournent vers des prestataires de niche comme Recapp ou Spitch, qui se spécialisent dans le sujet.

## 4.3 Résumé des possibilités d'utilisation

Le tableau ci-après résume les principales assertions quant à l'utilisation de systèmes de reconnaissance vocale au niveau judiciaire:

| Approprié | Pas (encore) approprié |
|---|---|
| **Administration**<br>• Courriers, e-mails<br>• Utilisation de textes et documents standard | **Situations avec plusieurs locuteurs**<br>• Auditions (dialogue)<br>• Audiences au tribunal |
| **Travail de chancellerie**<br>• Décisions, ordonnances, etc.<br>• Correspondance spécifique au cas | **Langues minoritaires / dialectes**<br>• réto-romanche<br>• suisse allemand<br>• accents prononcés |
| Partout où un dictaphone est utilisé aujourd'hui | Traductions de langues étrangères |

**Tableau 2: Possibilités d'utilisation dans le domaine judiciaire**

Il est certain que l'avantage principal de l'utilisation de systèmes Speech-to-Text dans le domaine judiciaire réside **actuellement** dans la saisie directe de textes dans les applications cibles.

## 4.4 Incidence sur les processus de travail

Comme indiqué dans le paragraphe précédent, certaines étapes de travail peuvent être simplifiées à l'aide de logiciels de dictée, par ex. la rédaction directe d'une décision par un juge. Cela signifie toutefois que certaines tâches courantes disparaissent alors pour les collaborateurs des services administratifs ou des secrétariats. En d'autres termes, des domaines d'activité peuvent ainsi être modifiés, ce qui est synonyme d'opportunités mais aussi de certains risques. Le Tableau 3 présente certains aspects correspondants:

| Opportunités | Risques |
|---|---|
| **Processus plus efficaces**<br>• Création directe du document<br>• Pas de transcription manuelle<br>• «Parler est plus rapide que taper à la machine» | Modifications de processus actuels qui fonctionnent bien |
| Amélioration de la qualité | Manque d'acceptation |
| Technologies et méthodes modernes | Avantages non perceptibles |
| **Suppression de matériel spécial**<br>• Dictaphones<br>• Pédales | Attentes erronées (envers le système) |
| **Adaptation des profils professionnels**<br>• Nouvelles tâches<br>• Davantage de responsabilités | Défis techniques |
| | Dépenses supplémentaires au cours de la phase d'introduction |

**Tableau 3: Opportunités et risques liés à l'utilisation de systèmes de reconnaissance vocale**

# 5 Acquérir et introduire une solution Speech-to-Text

## 5.1 Interrogations centrales

Dès lors qu'il est question d'acquérir un produit – en l'espèce une solution Speech-to-Text – une réponse doit être apportée à certaines interrogations. L'objectif visé est de rappeler les conditions-cadre en matière d'exploitation ainsi que les valeurs-clés envers le produit. Ces questions permettent ensuite de formuler les exigences fonctionnelles et non fonctionnelles. Les interrogations possibles à ce sujet sont précisées dans le tableau suivant:

| Questions liées aux exigences non fonctionnelles |
|---|
| Où la solution doit-elle ou est-elle censée être exploitée? (interne / externe) |
| Qui doit ou est censé gérer ou assurer la maintenance de la solution? (interne / externe) |
| Combien de personnes vont utiliser le système? |
| Quelle est la composition des groupes d'utilisateurs? |
| Combien de personnes vont utiliser le système? |
| Quels sont les aspects de protection des données à prendre en compte? |
| Quel est le cadre budgétaire? (budget) |
| Comment se présente le système dans son ensemble pour les utilisateurs? |
| Quelles sont les restrictions ou exigences pour les utilisateurs? |
| Quelles sont les prescriptions à respecter en ce qui concerne l'utilisation de produits logiciels? |
| **Questions liées aux exigences fonctionnelles** |
| Quelles sont les langues de transcription qui doivent être prises en charge? |
| Quelles fonctions doivent être proposées par la solution? (réponse à l'aide du Tableau 1) |
| Quels formats de données (intrant / extrant) doivent être pris en charge? |
| Quelles possibilités de saisie la solution doit-elle proposer? |
| Quelles possibilités de configuration la solution doit-elle proposer? |
| Quels problèmes concrets sont résolus ou amoindris par l'utilisation d'une solution Speech-to-Text? |

**Tableau 4: Sélection d'interrogations fondamentales pour l'acquisition du produit**

## 5.2 Planification de la procédure

Comme pour chaque projet, un plan d'action doit être défini pour l'acquisition et l'introduction d'un système de reconnaissance vocale. Il peut être divisé en deux phases principales: préparation ainsi qu'acquisition et mise en service.

### 5.2.1 Préparation

La phase de préparation débute par une analyse approfondie des exigences fonctionnelles et non fonctionnelles. Celles-ci peuvent être définies et précisées en fonction d'interrogations fondamentales (voir par ex. le ch. 5.1). Dès que les exigences ont été tirées au clair et enregistrées, l'évaluation d'un produit

approprié peut être lancée. Il convient pour cela d'évaluer les solutions possibles sur la base de critères prédéfinis (conformément aux exigences).

Une part considérable de la description des exigences est la définition des critères d'acceptation correspondants. Ces critères sont pris en compte, en plus de l'évaluation du produit, pour la description des cas à tester. Cette tâche doit impérativement être réalisée au cours de la phase de préparation.

Des aspects concernant la planification et la clarification du calendrier et des ressources, de même que la mise au point d'un plan de mise en œuvre (avec définition de la méthodologie de projet) forment une troisième part importante de la phase de préparation.

### 5.2.2 Acquisition et mise en service

L'installation du produit évalué (et acquis) marque le début de la mise en service. Le système Speech-to-Text est alors installé dans l'environnement cible ou les environnements cibles et configuré.

Après une formation des utilisateurs au niveau souhaité, une phase pilote devrait permettre de rassembler les premières expériences et de dégager les premiers enseignements. Il est recommandé de maintenir un cercle relativement restreint d'utilisateurs au début, afin que les retours puissent être traités et rapidement intégrés à l'optimisation et à la configuration. Le cas échéant, il convient également de préciser les exigences sur la base des informations obtenues.

Si la phase pilote se déroule correctement (décision positive quant à l'utilisation de la solution évaluée et testée), le déploiement peut alors être fait pour le cercle d'utilisateurs final. Ce déploiement devrait aussi se dérouler progressivement et par groupes de taille réaliste. Il faudra encore déterminer selon quels critères (langue, canton, office, etc.) il conviendra de procéder.

L'expérience montre qu'il faut prévoir suffisamment de temps pour le déploiement et la phase d'apprentissage. Une période de test de deux mois, par exemple, permet aux utilisateurs d'avoir suffisamment de temps pour découvrir les différentes facettes du système.

# 6 Recommandations

Au cours d'entretiens avec la direction de projet HIJP, il s'est avéré que l'intérêt pour les solutions Speech-to-Text était **particulièrement marqué pour les auditions et les audiences**. Comme expliqué, il n'y a pour l'heure pas encore de systèmes pouvant être utilisés dans la pratique. Il s'ensuit qu'il n'y a pour l'instant (encore) aucun sens à mettre en place un environnement de test pour le domaine HIJP.

Cependant, il est recommandé, compte tenu des activités de développement en matière de reconnaissance de plusieurs locuteurs et de transcription de dialectes, de **continuer à suivre**, ne serait-ce que de manière marginale, **l'évolution de la question** et de garder le contact avec des fournisseurs de solutions potentielles.

Il est en outre recommandé de **réfléchir aux exigences concrètes envers un système de reconnaissance vocale** et de les regrouper. A cet égard, des exigences fonctionnelles et non fonctionnelles jouent un rôle central pour la sélection d'une solution Speech-to-Text appropriée (cf. Tableau 1).