30. srpna 2022 | Jiří Špaček | jiri.spacek@rozhlas.cz | +420 725 594 388 | Český rozhlas

Syntéza hlasu - 2. fáze

Technická specifikace pro veřejnou zakázku Syntéza hlasu 2. fáze

Obsah

Slovník

- **ČRo** - Český rozhlas
- **Neurální hlas** - vysoce přirozeně znějící řeč syntetizovaná pomocí hlubokých neurálních sítí
- **Syntéza** - systém pro generování umělého neurálního hlasu
- **SSML** - značkovací jazyk na základě XML pro označování a konfiguraci
- **iR** - zpravodajský server iROZHLAS
- **SH** - Syntéza hlasu
- **VZ** – veřejná zakázka

Technická specifikace

1 Záměr ČRo

Český rozhlas má záměr zavést do své infrastruktury systém pro synteticky vytvářená audia s neurálními hlasy. Projekt by měl postupně sloužit pro servisní hlášení v aplikacích ČRo, načítání zpravodajských článků, načítání sumarizací a doporučení, načítání kratších i delších promluv do publicistický a případně i uměleckých pořadů a jiné využití v delším časovém horizontu.
Systém bude ČRo požadovat jako službu i dílo, tedy dodavatel by dodával administrační rozhraní pro administrátora pro nastavení systému, API pro předávání SSML, samotný systém - generátor audií a nové hlasy.
Projekt je rozdělen do tří fázi. V první fázi ČRo zjišťovalo možnosti a využití syntézy hlasu. V druhé fázi, která je předmětem této veřejné zakázky, hodlá zavést syntézu hlasu pro servisní a obsahové služby ve kvalitě neurálního hlasu. Ve třetí fázi ČRo uvažuje o systému s hlasy pro kreativní výrobu či syntetická média a vícejazyčné obsahy.

2 Syntéza hlasu 2. fáze

V druhé fázi předpokládáme zavedení robustního řešení automatizované produkce audií s umělým hlasem, který popisuje tato technická specifikace vycházející z POC první fáze.
Jde o zavedení on-premise řešení automatizovaného generování audií, on-demand generování audií s administrací nastavení výstupu a pro využití i v komponovaných pořadech. ČRo rozhodně zamýšlí vytváření nových hlasů ČRo, zavádění výslovnosti slov či případně zavedení anglického jazyka apod.
Cílem není zavedení procesu převedení textu na hlas, ale poskytovat našim čtenářům a posluchačům bezproblémový a příjemný poslech zajištěný automatizovaným systémem. Tedy kvalita výstupu je prioritou, nikoliv systém sám o sobě.

Veřejná zakázka bude hodnocena na základě nabídkové ceny a na základě ukázky syntézy zpravodajského textu ze serveru iROZHLAS, který pro potřeby hodnocení systém dodavatele online vygeneruje *(ČRo zašle požadavek – SSML soubor - např. na API a získá zpět link s audiem ke stažení)*. Obě položky, cena a kvalita výstupu, budou mít nejspíš stejnou váhu pro konečné hodnocení.

Diagram workflow systému naleznete pod bodem 7 Návrh systému.

3 Obvyklé use-case využití syntézy

1. Redaktor publikuje článek a tím vytváří požadavek na generování syntézy pro FE dle zvolené šablony
2. Redaktor si vyžádá zkušební syntézu článku pro kontrolní poslech
3. Redaktor si vyžádá opravu (provedl úpravy) článku.
4. Uživatel systémů ČRo si vyžádá ad hoc syntézu do svého audio příspěvku dle nastavení svého nastavení. Tedy zašle SSML z našeho generátoru.
5. Aplikace (např. mobilní) si vyžádají vytvoření ad hoc syntézy pro servisní účely

4 Přehled požadavků

4.1 Základní funkční požadavky

Jednotlivé funkční požadavky jsou rozepsány podrobně v kapitole 5.

1. Generátor výstupního audia v požadovaném formátu a kvalitě
2. Konfigurátor syntézy pro nastavení možností a rozsahu kvalit
    a. SSML tagy s parametry a rozsahem hodnot
    b. nové hlasy
    c. výslovnost atd.
3. API
    a. Zadávání on-demand požadavků na výrobu audií
    b. Zadávání požadavků na konfigurátor hlasu
4. Pre-processing
    a. normalizace SSML
    b. SSML validátor
5. Plánovač úloh, fronty
6. Post-Processing
    a. Vkládání externího audia (wav, mp3) do syntézy
    b. Výsledný audiomix
    c. Efekty
7. Logování požadavků a procesů (Log)
8. Výroba neurálních hlasů pro ČRo na základě dodaných podkladů

4.2 Základní ne-funkční požadavky

Systém bude řešený on-premise. Měl by splňovat:

1. Výkon syntézy
    a. Zpravodajský server ca 150 požadavků denně v rozmezí 06.00-24:00 v různých kvalitách, jeden požadavek ca 4 minuty, celkem ca 10 hodin
    b. Ad-hoc požadavky v jednotkách denně.
    c. Noční časy budou využívány pro dogenerování audií ke starší článkům.
2. Dostupnost a spolehlivost
    a. SW je optimalizovaný pro provoz v režimu 24 / 7 s vysokou dostupností (bude řešeno v rámci serverovny IT ČRo)

    b. pro údržbu SW je písemně dohodnuto servisní okno v maximálním rozsahu 4 hodin měsíčně v době, kdy nejméně omezuje potřeby Objednatele
3. Rozšiřitelnost
    a. O další moduly pro zlepšování kvality výstupů (produkční i postprodukční)
    b. Nové hlasy
4. Škálovatelnost systému
    a. Možnost škálování systému bez nutnosti dočasného vyřazení z provozu.
    b. Možnost rozšíření o nové funkcionality, hlasy a jejich kvality bez nutnosti dočasného vyřazení z provozu.
5. Integrace
    a. Vstup a výstupy napojené na systémy ČRo - webové a mobilní aplikace
    b. API na straně systému syntézy pro zadávání požadavků a konfigurace
6. Licence
    a. Nevýhradní licence
    b. Na dobu neurčitou
    c. Využití systému pro ČRo a subjekty spolupracující s ČRo včetně komerčního využití
    d. Licence na použití syntetického hlasu
    e. Licence na zpracování hlasu pro syntézu
7. Ladění hlasů
   Dle požadavků ČRo a ve spolupráci se zástupci ČRo proškolit pracovníky a do praxe zavést konfigurace hlasů (šablony), které je možné používat dle potřeb ČRo.
8. Bezpečnost
   Data zadavatele, která by byla přístupná mimo síť ČRo, musím být chráněna proti zneužití.
9. Právní ochrana
   Zadavatelem poskytnutá data zůstávají v jeho držení.
10. Dočasné HW řešení
    Dodavatel poskytne ČRo dočasné HW řešení pro testování a případný vývoj systému před nasazením na produkční HW zadavatele.
11. Dočasný neurální hlas
    Dodavatel poskytne po dobu nezbytně nutnou vlastní neurální hlas pro testování systému.

4.3 Servis a podpora - SLA

Servis a Podpora se týkají programového vybavení nikoliv HW, který bude ve správě ČRo. Jde tedy o update a upgrade SW, chyby, vyřizování change requestů.

Servisní doba Dodavatele je v každý pracovní den v režimu 8 hodin / 5 dnů v týdnu.

Je dodržována reakční lhůta (fyzickým člověkem, ne automatem) a lhůta pro odstranění vady od nahlášení závady dle následující tabulky:

Stupeň priority závady	*Popis závady*	*Reakční lhůta od oznámení požadavku (v rámci režimu 8/5)*	*Lhůta pro odstranění vady od oznámení požadavku*
1- Kritický incident	více než 20% požadavků není odbaveno, systém je nefunkční	2 hodiny	1 pracovní den
2 - Vážný incident	požadavky jsou z 80 % odbaveny, není funkční administrace systému	4 hodiny	2 pracovní dny
3 - Běžný incident	závada umožňující práci systému s pomocí náhradního pracovního postupu	1 pracovní den	3 pracovní dny
4 - Běžný požadavek	např. úprava konfigurace nebo drobná chyba, která neovlivňuje činnost	2 pracovní dny	5 pracovních dnů

Pro změnové požadavky (nové funkcionality apod) dodá dodavatel odhad pracnosti nejpozději do 15 pracovních dnů.

4.4. Seznam technických požadavků

Technická specifikace serverů na straně ČRo. Případné další požadavky dodavatel upřesní:

Počet serverů k dispozici:
2 v mirroru

Server - konfigurace:

- OS
    - LinuxRed Hat / CentOS / Ubuntu / Windows Server
- Počet serverů:
    - 2 paralelně pro případ výpadku
    - active / active
- Bez nutnosti zakoupení dalšího licencovatelného SW
- Možný přístup do SQL
- Bez grafické karty

Kód / PN	Popis nabízeného produktu/služby	Počet
	Intel varianta pozn. uváděné ceny jsou doporučené koncové	
P05172-B21	HPE ProLiant DL380 Gen10 Plus 8SFF NC Configure-to-order Server	1
P05172-B21 B19	HPE DL380 Gen10 ICX CTO Mod-X 8SFF	1
P36933-B21	Intel Xeon-Gold 6334 3.6GHz 8-core 165W Processor for HPE	2
P06029-B21	HPE 16GB (1x16GB) Single Rank x4 DDR4-3200 CAS-22-22-22 Registered Smart Memory Kit	12
P27194-B21	HPE ProLiant DL300 Gen10 Plus 2U 8SFF x1 Tri-Mode 24G U.3 BC Front Drive Cage Kit	1
P28586-B21	HPE 1.2TB SAS 12G Mission Critical 10K SFF BC 3-year Warranty Multi Vendor HDD	2
P37038-B21	HPE ProLiant DL380 Gen10 Plus x8/x16/x8 Primary FIO Riser Kit	1
P26253-B21	Broadcom BCM57416 Ethernet 10Gb 2-port BASE-T Adapter for HPE	1
P26325-B21	Broadcom MegaRAID MR216i-a x16 Lanes without Cache NVMe/SAS 12G Controller for HPE Gen10 Plus	1
P08449-B21	Intel I350-T4 Ethernet 1Gb 4-port BASE-T OCP3 Adapter for HPE	1
P37042-B21	HPE ProLiant DL300 Gen10 Plus 2U Standard Fan Kit	1
P38995-B21	HPE 800W Flex Slot Platinum Hot Plug Low Halogen Power Supply Kit	2
BD505A	HPE iLO Advanced 1-server License with 3yr Support on iLO Licensed Features	1
P13771-B21	HPE Trusted Platform Module 2.0 Gen10 Plus Black Rivets Kit	1
873763-B21	HPE DL38X Gen10 8 SFF Front Cage Removal FIO Option	1
P22018-B21	HPE DL38X Gen10 Plus 2U SFF Easy Install Rail Kit	1
P22020-B21	HPE DL38X Gen10 Plus 2U Cable Management Arm for Rail Kit	1
P27095-B21	HPE ProLiant DL380 Gen10 Plus High Performance Heat Sink Kit	2
HU4B2A5	HPE 5Y Tech Care Basic Service	1
HU4B2A5 ZSB	HPE Proliant DL380 Gen10+ Support	1
HU4B2A5 R2M	HPE iLO Advanced Non Blade Support	1

5 Podrobný popis funkčních požadavků

5.1 Systém pro generování audia se syntetickým hlasem

Systém získává data v SSML z API a generuje na základě zadané konfigurace syntézu hlasu do specifikovaného audiosouboru. V případě paralelního zpracování provádí finalizace do jednoho souboru bez ztráty kvality a skrze API vystavuje link ke stažení.

5.1.1 Škálování generování audií

Možnost paralelního zpracování jednoho textu na více procesech, tzn. systém umí rozdělit delší texty na více kratších částí.

5.1.2 Délky zpracovávaných textů

Od 1 slova až ca 30000 znaků dlouhé články (extrém).
Naprostá většina textů je v délce 2000-3500 znaků.

5.1.3 Délka procesu generování audiosouboru

ČRo vyžaduje near-time a off-line generování hlasu.
Generování v nejlepší kvalitě může být generováno nejpomaleji 1:1 v čase běžné promluvy a to bez rozložení na více procesů. Při paralelním zpracování bude čas úměrně kratší dle množství použitých procesů současně.

5.1.4 Hlasy externích dodavatelů

Systém bude připraven na možnost využívat v budoucnu i hlasy od jiných dodavatel viz diagram workflow systému

5.1.5 Audio formát a zasílání / ukládání generovaného souboru

- Typy výstupního audio souboru: AAC, MP3, WAV
- Vzorkovací frekvence (Sample-rate): 48 kHz default
- Bitová hloubka (Bit-depth): 16 / 24
- Výstup: Stereo i když jde o mono

5.1.6 Vývojové prostředí systému

Dodavatel provádí vývoj ve svých obvyklých frameworks a dle svých obvyklých postupů. Zadavatel očekává standardizované postupy ve vývoji a testování.

5.1.7 Testování výslovnosti

Uživatel má možnost otestovat výslovnost slova zasláním vybraného slova na test (např. pro zjištění jak čte systém jméno “Heidegger” pro případný fonetický zápis).
V případě, že text je příliš dlouhý, systém si sám v preprocessingu text rozdělí na kratší části a zařadí do fronty ke zpracování tak, aby výsledkem bylo jeden audio soubor (viz audiomix).

5.2 Konfigurátor systému syntézy

Konfigurátor je aplikace s vlastním UI, která nastavuje systém generátoru syntézy.
Tedy dostupné hlasy, přípustné SSML tagy s atributy, nová slova s výslovností a výslovnosti.

Nastavení bude uložené v DB ČRo a sdílené s konfigurátorem SSML, který je vyvíjen na straně ČRo.

5.2.1 Zavedené SSML Elementy s parametry

SSML: parametry vycházejí z konvence W3.org, jejich použití je rozepsáno v jednotlivých položkách nastavování kvality v konfigurátoru. V základu jde o tyto elementy, tučně povinné, ostatní k diskusi.

- audio
- **break**
- **emphasis**
- lang
- lexicon
- lookup
- mark
- meta
- metadata
- **p**
- **phoneme**
- **prosody**
- **say-as**
- sub
- **s**
- token
- **voice**
- w

Podrobný popis úpravy kvality neurální hlasu elementy a jejich atributy v SSML.
Dodavatel může nabízet vlastní atributy nebo elementy pro plné využití kapacit generátoru syntetického hlasu.

Popis povinných elementy:

1. Odstavec
    a. element: **p**
    b. označení začátku a konce odstavce
2. Věta
    a. element: **s**
    b. označení začátku a konce věty
3. Prozódie
    a. element: **prosody**
    b. nastavení kvalit promluvy
    c. atributy:
        i. volume
            - hlasitost, ruční nebo předvolená
        ii. pitch **(není povinný atribut pro přihlášení do VZ)**
            - výška / hloubka generovaného hlasu
        iii. rate
            - rychlost
        iv. contour **(není povinný atribut pro přihlášení do VZ)**
            - průběh intonace
4. Fonetická výslovnost
    a. element: **phoneme**
    b. fonetický přepis pro nové či neobvyklé výrazy
5. Nahrazení slova známým slovem

    a. element: **sub**
    b. zapsání vlastního výrazu (který systém zná) či rozepsání zkratky
6. Zavedení šablony výslovnosti
    a. element: **say-as**
    b. generování dle jiného výrazu
7. Šablony pro standardní styl výslovnosti
    a. element: **say-as** s předvolenými atributy pro formát
    b. např. šablony pro: čísla a procenta, měny, e-maily a www stránky, čas, datum, telefonní číslo či matematické znaky, politické strany, instituce, jména lidí
8. Hlas
    a. element: **voice**
    b. výběr z dostupných neurálních hlasů
9. Pauza
    a. element: **break**
    b. Standardní i volitelná délka pauzy v ms

Popis některých nepovinných a zamýšlených elementů

10. Morfing
    a. element: voice
    b. atribut: morph
        i. varianty dle potřeb
11. Emoce
    a. element: dle dodavatele
    b. varianty: strach, radost, smutek, štěstí, neštěstí, pláč, smích
12. Úprava gender
    a. element: dle dodavatele
    b. varianty: muž, žena, neurčitý (gender neutral)

13. Úprava věku
    a. element: dle dodavatele
    b. varianty: dítě, teenager, mladý člověk, střední věk, senior
14. Styl řeči podle typu obsahu (jako šablona)
    a. Lze řešit na straně generátoru SSML
    b. element: dle dodavatele
    c. varianty:
        i. článek iR
        ii. publicistický audiopořad
        iii. citace
        iv. servisní hlas
        v. custom
        vi. default
15. Šablona
    a. element: dle dodavatele
    b. varianty: komentář, rozhovor, zpráva...

5.2.2 Nastavení výslovnosti

Systém umožňuje nastavit výslovnost slov a to na základě elementu say-as.

5.2.3 Nová slova

Systém umožňuje ruční zavádění nových slov i s jejich výslovností, tuto službu zavádění slov poskytuje dodavatel.
V případě že v SSML je nový výraz včetně výslovnosti, měl by být do knihovny zaveden pro schválení.
Schvalování by mělo probíhat skrze dodávané API.

5.2.4 Nové hlasy

Systém umožňuje zavedení nového hlasu a to jak interního tak externího, tedy mimo síť ČRo za firewallem, viz diagram systému bod 7.

5.2.5 Kalibrace hlasitosti

nastavení defaultní hlasitosti generovaného audia pro finální mastering, pokud není v požadavku uvedeno jinak.

5.2.6 Další nastavení

- Maximální délka textu pro rozdělení na kratší části
- Rozsah priorit pro zařazení do fronty ke zpracování (5.4.2.2)
- Dostupné uživatelské slovníky
- Nastavení validátoru

5.3 API

Typ: API RESTful
Framework: např. Swagger pro snadnou dokumentaci a testování
Dokumentace: v rámci URL API
Umístění: servery ČRo
Přístupnost: v rámci sítě ČRo bez autentizace, mimo síť ČRo s autentizací

5.3.1 Základní funkce API:

- spouštění procesů
    - metody: PUT, GET, POST
- zadávání úkolů
    - metody: PUT, GET, POST
- získávání výsledků
    - metoda GET

Procesy provádějí generování audií.
Úkoly jsou SSML soubory ke zpracování.
Výsledek může být link na audiosobor ke stažení.

5.3.2 Sekundární funkce:

5.3.2.1 Zasílání SSML k validaci
5.3.2.2 Odeslání na preprocessing
5.3.2.3 Vkládní / potvrzování nových slov
5.3.2.4 Ovládací prvky pro SSML
5.3.2.5 Testovací verze SSML (kontrolní poslech článku)

5.4 Pre-processing SSML

Pre-procesingem je zamýšleno přijetí SSML souboru skrze API, jeho validace a normalizace textu a zasílaného souboru SSML pro potřeby jeho zpracování generátorem syntézy.

5.4.1 Validace výslovnosti

Kontrola výskytu výrazů ve slovníku a zaslání nesrovnalostí nebo stavu OK na API pro ČRo. ČRo bude zprávy přijímat a zavádět do seznamu pro ruční kontrolu a případné ruční zásahy ve slovnících.

5.4.2 Validace SSML

Kontrola platnosti kódování a normalizace kódu a s případným hlášením změn.

5.4.3 Normalizace SSML pro zpracování generátorem syntézy

1. Převádí čísla, zkratky, datumy, politické strany, instituce, na celá slova
    a. Systém si musí pamatovat co se jak vyslovuje
    b. Systém by si měl automaticky odvodit i pády
2. Možnost zadat fonetickou transkripci k libovolnému slovu
    a. Zadává se česky
    b. Trvale nebo ad hoc
3. Výslovnost
    a. Možnost zaslat vlastní slovo s vlastní výslovností - zapsat česky foneticky - jména lidí, státy, instituce, dokumenty
    b. Možnost změnit výslovnost zavedenou v systému
    c. Možnost změnit čtení v preprocesingu defaultně zadané výslovnosti
4. Rozdělení textu podle spuštěných procesů tak aby proces zpracování mohl probíhat paralelně na více jádrech nebo procesorech
5. Automatická šablona dle zdroje podkladů
6. Validace SSML
7. Rozdělení na kapitoly - přidání značek kapitol pro rychlý přesun - základ bude v SSML (zavést do šablony - vkládat značky dle H2)

5.5 Plánovač fronty

5.5.1 Řazení úkolů do fronty

Vytvoření fronty požadavků a jejich správa podle priorit.
Rozčlenění požadavků na interní a externí požadavky.

5.5.2 Prioritizace požadavků

Možnost prioritizace požadavků ve frontě její administrace. Je možné zavést priority ke zpracování defaultně dle číselníku (1-10) a zařazovány do fronty budou dle hodnoty priority preprocessingu / před zasláním na API. Aktuálně řadit dle pravidel:

1. Aktualizace publikovaného článku (2)
2. Zpravodajský článek k publikaci (4)
3. On-demand požadavek s vyšší prioritou (6)
4. Zpravodajský článek k testovacímu poslechu (8)
5. On-demand požadavek s nižší prioritou (10)

5.5.3 Hlasy externích dodavatelů

Příprava pro potenciální generování syntézy u externího dodavatele, tedy zaslání SSML s tokenem nebo jiným přístupovým kódem na API třetí strany a získání zpět audia (URL + stažení a další zprocesování).

5.6 Post-processing

Kompletace včetně generování finálního audia dle požadavků v SSML.
Systém generování syntetického hlasu musí umožnit vložení externího audia.

1. Výsledný mix
2. Mastering
3. Efekty
4. Verzování

5.6.1 Výsledný mix

Systém sestaví finální audio a připraví pro mastering. Kompletace vychází z požadavků verzí zadaných v SSML a dle rozdělení úkolu na více segmentů a externích služeb či zdrojů.

- Spojení více audií v jeden výstup v případě paralelního generování dle SSML
- Zavedení externích výstupů,
    - externí hlas,
    - externího zvuku / ruchu z katalogu,
    - externí audio
- *Zavedení úvodní / konečné znělky (rozvojová položka, není povinná)*
- *Zavedení jinglu (rozvojová položka, není povinná)*
- *Možnost zavedení podkresu (rozvojová položka, není povinná)*
- *Zavedení úvodní nebo závěrečné upoutávky (rozvojová položka, není povinná)*

5.6.2 Mastering vloženého audia a audiomix

Audio musí projít “masteringem” aby bylo ve stejném formátu jako hlas.

- Systém by měl zvládnout sám, downsample originálu ke kvalitě syntetického hlasu + možnost převést na stejný mono/stereo

- Výstup stereo (i když hlas bude mono)
- Přístup na zdroj audia
- Kalibrace hlasitosti
- Formát audia - možnost si změnit výstup na jiný (lepší - finální / horší - korektura)
- Fade in / fade out dle SSML
- Mix dle SSML včetně zavedení podkresu.

5.6.3 Efekty

Dostupné dle možností. Zvukové efekty segmentů.
Popis v šabloně SSML, v případě nutnosti zavedení vlastních elementů či atributů k elementům pro splnění požadovaného výsledku.

5.6.4 Verzování

Možnost generovat více verzí v postprocesingu (čistá verze, verze s jingly a hudbou, verze s přílepkem pro třetí strany apod.) vždy dle SSML.

U některé efektů nebo úprav je ekonomicky i technicky jednodušší provádět je až po dokončení

5.7 Logování procesů

Pro snadnou diagnostiku procesů systému požadujeme z počátku velmi podrobné logování procesů přístupné i vývojářům ČRo, které bude po zavedení zjednoušeno.

5.8 Šablony SSML

Šablony SSML jsou přednastavené košilky při generování SSML, např. Zpráva, Rozhovor, Sport, Long read apod. Spolupráce na přípravě šablon dodavatelem pro preprocessing dat. Např. varianty zpráv apod. Šablony budou vytvářeny mimo systém SH a to v nastavení SSML generátoru.

Generátor SSML zajišťuje ČRo a jeho výstupem je SSML soubor zaslaný na API SH dodané dodavatelem.

5.9 Neurální hlasy

5.9.1 Přechodné období

Dodavatel poskytne na dobu přechodnou vlastní neurální hlas pro potřeby testování a ladění systému a pro případný start než bude zaveden první neurální hlas pro Český rozhlas.

5.9.2 Výroba hlasů pro ČRo

- ČRo poskytne dodavateli nahrávky hlasu ve studiové kvalitě s i texty.
    - Dodavatel může poskytnout vlastní know how pro přípravu podlkadů k výrobě neurálního hlasu.
- Dodavatel na dodaných podkladech natrénuje neurální hlas a provede testování.

- Následně poskytne hlas k testování zadavateli a po jeho schválení zavede hlas do systému.
- ČRo bude zadávat podklady k výrobě dalších hlasů na základě dílčí smlouvy a sjednané ceny.
- Minimálně jeden hlas musí být vždy dostupný z interních zdrojů.
- Předpokládaný postup: Analyzátor textu -> neuronový akustický model -> neuronový vocoder

5.9.3 Externí hlasy

Dodavatel může z technických či časových důvodů doplnit sadu hlasů i externími hlasy viz bod 7.

5.9.4 Skiny hlasů

Dodavatel by měl navrhnout vlastní řešení jakým způsobem dále rozšiřovat možnosti neurálního hlasu ať novým přístupem při jeho generování či postprodukčně.

6 Další rozvoj, testování, SLA

6.1. Testování

Každá úprava, nová funkce nebo modul musí projít testováním na straně dodavatele.
Na stage / produkci mohou být postoupeny pouze části kódu, který prošel testováním bez výhrad a nebo za souhlasu zadavatele.

6.2. SLA

S dodavatelem bude do 6 měsíců od prvního spuštění uzavřena SLA na 1MD měsíčně na údržbu a drobné úpravy systému.

6.3 Další rozvoj

Další rozvoj bude vycházet z roadmapy projektu Syntéza hlasu, aktuálních potřeb na změny nebo další vývoj.
Požadavky budou specifikovány zvláště objednávkou nebo dílčí smlouvou.

Zejména se jedná:

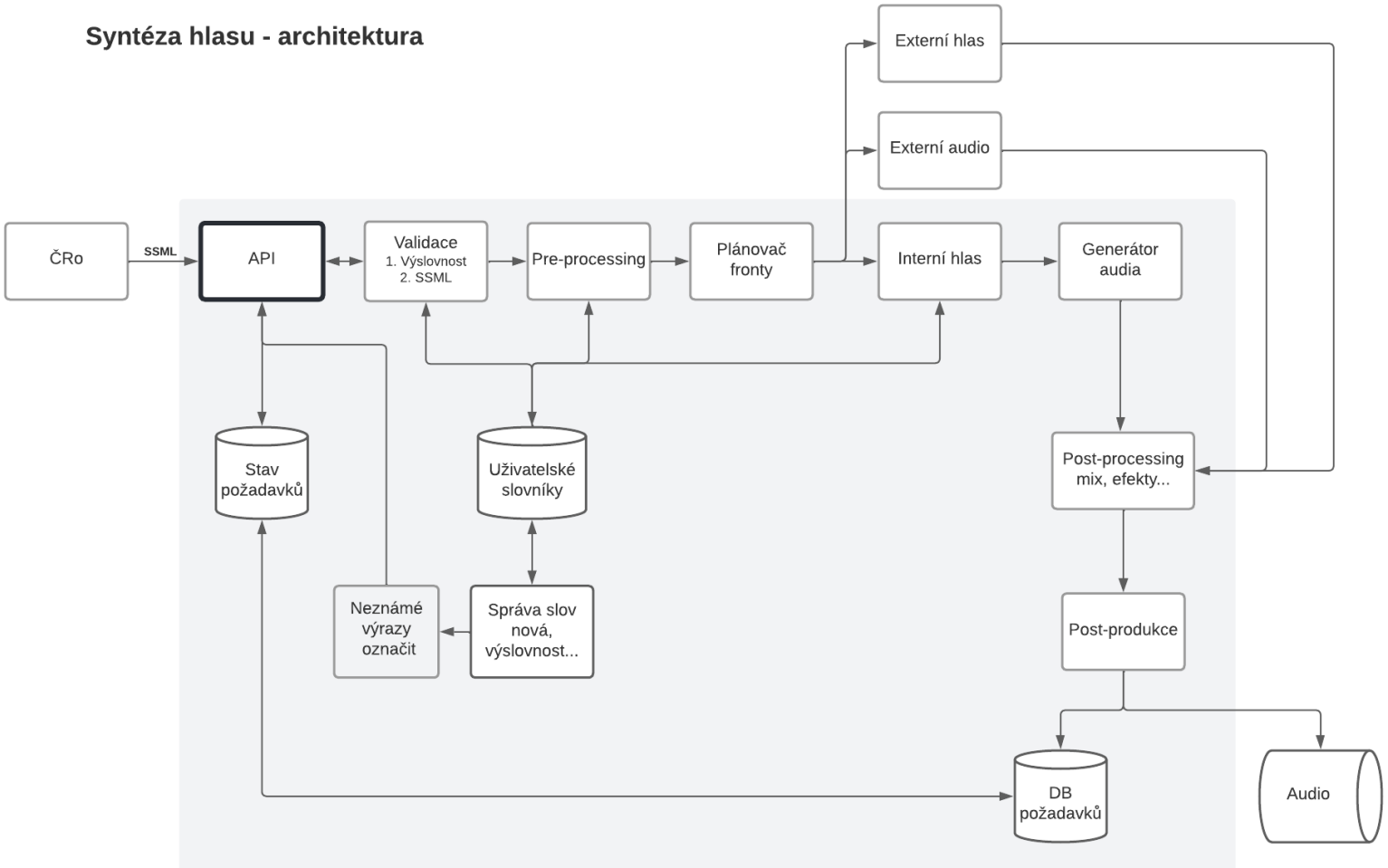
- Rozšíření portfolia syntetických hlasů
- Rozvoj možností úprav hlasu (emoce, věk, styl)
- Postprocesingové úpravy a efekty
- Využití pro třetí strany
- Spolupráce na vývoji kreativního studia
- Spolupráce na vývoji syntetického média
- Spolupráce na rozvoji hlasových služeb v aplikacích ČRo

7 Návrh systému - diagram

Digitální verze diagramu zde.

7.1 Stručný popis

1. ČRo vytvoří SSML soubor a zašle požadavek na API.
2. Na straně API dojde k validaci výslovnosti a SSML a přípravy podkladů ke generování audia i na základě uživatelských slovníků.
3. Evidovaný požadavek bude zařazen do fronty.
4. Na základě SSML bude zvolen generátor audia a následně bude audio vygenerováno a zmixováno dle nastavených požadavků.
5. Výstupem je audio ke stažení / použití v ČRo.



8 Předpokládaný harmonogram

Obecně předpokládáme, že od účinnosti smlouvy bude systém spuštěn do 3 měsíců a odladěn nejpozději do 6 měsíců.
Neurální hlas z podkladů ČRo by měl být dodán do ca 2 měsíců od předání podkladů.

	Týden															
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Smlouva	schválení, podpisy, registr															
Neurální hlas ČRo			předání podkladů							dodání						
Vývoj API			zahájení prací							dodání						
Zavedení systému			instalace, testování													
Hlas na přechodné období																
Produkční verze final																